Non-coding RNA and transcriptional regulation in CD4 T cell lineages

Catherine M. Evans

A thesis submitted to UCL for the degree of Doctor of Philosophy

Supervisors:

Dr. Richard G. Jenner

Prof. Michael R. Ehrenstein

Declaration

I certify that this thesis does not incorporate any material previously submitted for degree or diploma in any university and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Abstract

CD4 T cell lineage choice epitomises the ability of the immune system to become tailored to a specific threat and provides a framework for understanding the mechanisms behind cell specification. The differentiation of T effectors from naïve cells gives rise to pro-inflammatory lineages including T helper 1 (Th1) and Th2 and anti-inflammatory regulatory T cells (Treg). An additional lineage of Treg also exits the thymus in parallel to naïve T cells and together these Treg are required for prevention of autoimmunity. These T cell lineages are distinct in terms of their cytokine production and functional effects but also through their differences in gene expression and its regulation, which are orchestrated by the presence of lineage-specifying transcription factors specific for each lineage. In addition, post-translational modification of histones also provide insights into this transcriptional regulation and more recently the pervasive and tissue-specific transcription of multiple classes of RNA species without protein coding capacity, non-coding RNA (ncRNA), has been found to play a role in cell differentiation and function.

In this thesis I identify several ncRNAs with differential expression different T cell lineages. This includes ncRNAs upregulated Treg compared to T responders. The characterization of these, including their expression in the autoimmune context of systemic lupus erythematosus (SLE), is presented and their possible biological functions are examined. The relevance of histone modifications for influencing Treg identity in SLE is also investigated. An additional class of ncRNAs that originate from gene enhancer regions, eRNA, is also investigated in the context of Th1 versus

Th2 lineage choice. This enhancer transcription is increased genome-wide in Th1 cells at enhancers with high density T-bet binding in, termed ‘super-enhancers’. The functional relevance of these eRNAs, including at the super-enhancer upstream of the Th1 signature cytokine gene, IFNG, is also investigated in knockdown experiments.

Acknowledgements

First and foremost I thank my supervisors, Mike Ehrenstein and Richard Jenner, in making this thesis possible and for their guidance. Richard, your enthusiasm for the scientific process has been encouraging. I'm glad I learnt how to do science from you.

This thesis could not exist without the funding bodies, the Nuffield Foundation Oliver Bird Rheumatism Program and Lupus UK, and without David Isenberg for securing the Oliver Bird funding for UCL. You have been inspiring through your tireless involvement in your work, your cheery demeanour at all times and as the front man of the world's most successful (probably) rock/rheumatology-based-spoof band, *Lupus Dave and the Davettes*. I hope I have your energy when I am that close to retirement age! I would also like to thank the Oliver Bird organisational staff, particularly Vicki Hughes, who have worked hard to ensure us PhD students get the best personal and professional development we can. The annual conferences have been refreshing, focussing and rewarding.

Many lab members need thanking for their scientific and personal support; From Richard's lab: Arnulf Hertweck, Lenka Skalska, Alessandro Riccombeni, Filipa Reis, Manual Beltran-Nebot, Keijo Viiri, Aditi Kanhere and Chris Yates; From Mike's lab: Helen Baldwin, Clare Notley, Shawn Ellis, Graham Wright, Dao Nguyen, Rachel Byng-Maddick, Jenny McGovern, Amara Ezeonyeji and Sandra Martins. Special mentions go to other scientists who have taken their time unnecessarily to aid me. This includes Laurie Menger, Tan Choon Ping, Adam Fletcher and Kristine Oleinika. There are many others who have provided useful conversations

or reagents. No scientist is an island! Support from all these people, at the expense of their own time, has been a comfort. In addition many blood donors need thanking, both healthy ones and those who have attended the UCH Rheumatology clinic. This work has relied on them.

I wouldn’t be writing this if an enthusiasm for science, the scientific process and a passion for truth and knowledge about how the world works hadn’t been sparked by many teachers, both professionally and informally. These include Dad, Mrs Piper, Mrs Shuttleworth, Mr Hill and Rosemary Gale. Thank you, Stuart, for so much support in so many ways that has really helped me, not least your ability to always see the bright side and bring a smile to my face no matter what. Finally, thank you to my family for providing me with stability, support and care throughout my life.

Contents

Figures

Tables

Abbreviations

AhR	Aryl hydrocarbon receptor
APC	Antigen presenting cell
APC	Antigen presenting cell
ATP	Adenosine triphosphate
atRA	All-trans retinoic acid
bp	Base pair
cAMP	Cyclic adenosine monophosphate
CCR	Cysteine-Cysteine amino terminal motif chemokine receptor
CD	Cluster of differentiation (suffixed by idenfication number)
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation coupled to sequencing
CNS	Conserved non-coding sequence
CTP	Cytosine triphosphate
DAPI	Diamidino phenylindole
DC	Dendritic cell
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucelotide triphosphate
EAE	Experimental autoimmune encephalomyelitis
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial-to-mesenchymal transition
eRNA	Enhancer RNA
EST	Expressed sequence tag
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
GALT	Gut-associated lymphoid tissue
GO	Gene ontology
GTP	Guanosine triphosphate
H3	Histone 3
HC	Healthy control
HOT	Highly occupied target
ID	Identification
Ig	Immunoglobulin
IL-	Interleukin (suffixed by identification number)
iTreg	induced Treg
Kb	Kilobase
LincRNA	Long intergenic ncRNA
LNA	Locked nucleic acid
lncRNA	Long ncRNA
MACS	Magnetic associated cell sorting; Model based analysis for ChIP-Seq
MHC	Major histocompatibility complex
miRNA	MicroRNA
mRNA	Messenger RNA
NAT	Natural antisense transcript
ncRNA	NcRNA

ncRNA-a	Activating ncRNA
NK	Natural killer
NKT	Natural Killer T cell
nt	Nucleotide
NTP	Nucleoside triphosphate
nTreg	Naturally occurring Treg
P-TEFb	Positive transcriptional elongation factor b
PBMCs	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PCA	Principle component analysis
PCR	Polymerase chain reaction
pfp	predicted false positive
PMA	phorbol myristate acetate
PRC2	Polycomb repressive complex 2
prompts	Promotor upstream transcripts
QPCR	Quantitative polymerase chain reaction
RA	Rheumatoid arthritis
RIN	RNA integrity number
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
RNAi	RNA interference
RNAPII	RNA polymerase II
ROSE	Rank Ordering of Super-enhancers
RPKM	Reads per killobase per million sequenced
RPMI	Roswell Park Memorial Institute
rRNA	Ribosomal RNA
SEM	Standard error of the mean
siRNA	Short interfering RNA
SLE	System lupus erythematosus
snoRNA	Short nucleolar RNA
SNP	Single nucleotide polymorphism
snRNA	Short nuclear RNA
STAT	Signal transducer and activator of transcription
SV40	Simian vacuolating virus 40
TAP	Tobacco acid pyrophosphatase
TCDD	2, 3, 7, 8-Tetrachlorodibenzo-ρ-diooxin
TCR	T cell receptor
TCR	T cell receptor
TES	Transcription end site
TGF-β	Transforming growth factor beta
Th1	T helper 1
Th17	T helper 17
Th2	T helper 2
TIP	Transcription initiation platform
Tr1	T regulatory 1 cell
Tr35	T regulatory 35 cell
Treg	Regulatory T cell
Tresp	Responder T cell

tRNA	Transfer RNA
TSDR	Treg specific demethylated region
TSS	Transcription start site
TTP	Thymidine triphosphate
TUCP	Transcripts of uncertain coding potential
UCLH	University College London Hospital
UCSC	University of California, Santa Cruz

1 Introduction

1.1 CD4 T cell lineages

CD4 T cell lineage choice epitomises the ability of the immune system to become tailored to a specific threat and additionally provides a framework for understanding the mechanisms behind cell specification. The identification of distinct subtypes of CD4+ T cells began with the observation that this population appeared heterogeneous in activity, particularly regarding B cell help [1]. The first two identified lineages, termed T helper 1 (Th1) and T helper 2 (Th2), were further defined by differences in cytokine production, with Th1 cells producing IFNγ and Th2 cells producing IL-4 [2], IL-5 [3] and IL-13 [4, 5] (Figure 1.1). The concept of distinct lineages was reinforced by the finding that these differences were stable during passage [2], demonstrating this specialisation to be heritable. The production of these signature cytokines underpins the functional importance of these specialised T cell lineages within the immune system. IFNγ secreted from Th1 cells activates macrophages [6] allowing clearance of intracellular pathogens, such as viruses and intracellular bacteria in what is termed a Type I response, while Th2 cytokines IL-4 and IL-5 support expansion of B cells and subsequent antibody production by plasma cells in the humoral immune response [7], leading to clearance of extracellular pathogens, such as extracellular bacteria and helminths [8] in a Type II response. This is the basis of the Th1/2 paradigm and is illustrated in Figure 1.1.

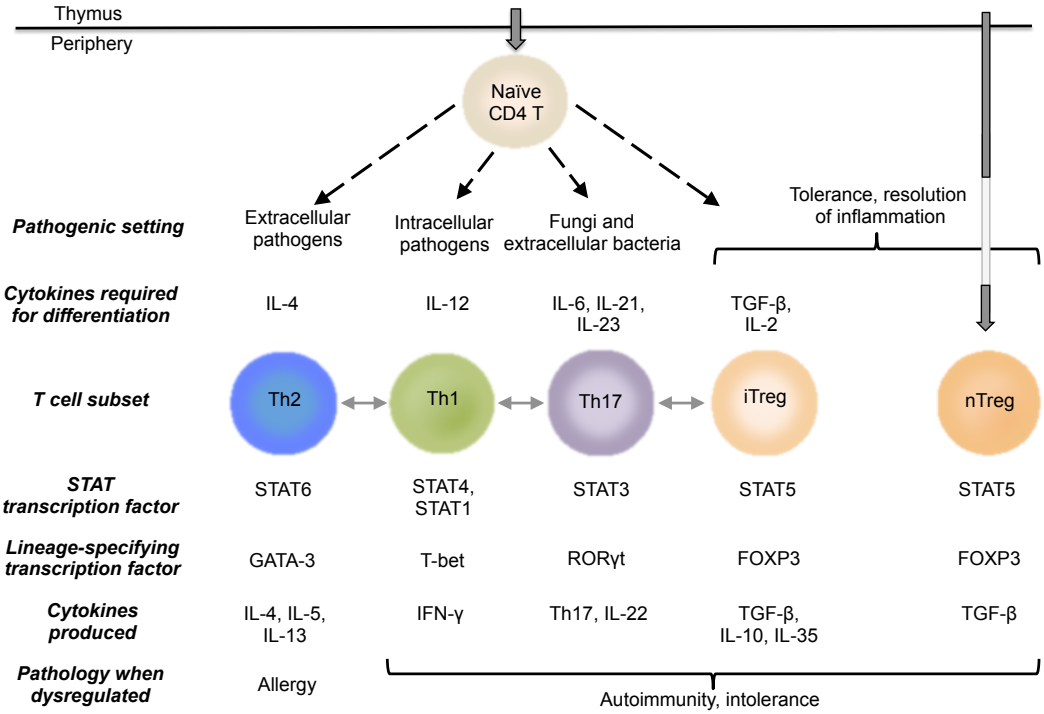


Figure 1.1: T cell differentiation

CD4 T cell lineage differentiation allows a response from the immune system to be tailored to the pathogenic setting and also provides a framework for understanding the mechanisms behind cell specification. Naïve CD4 T cells exit the thymus and upon stimulation differentiate into specialised subsets under the influence of cytokines that reflect the particular pathogenic threat. Signalling downstream of these cytokines leads to activation of STAT transcription factors which prepare the chromatin landscape for the expression of lineage-specifying transcription factors. In addition, naturally occurring Treg exit the thymus as a distinct lineage in parallel with naïve T responder cells. Each subset is characterised by the production of signature cytokines. There is some plasticity between these subsets, as illustrated by the grey arrows. Dysregulation of differentiation leads to immune pathology. STAT: signal transducer and activator of transcription; iTreg: induced regulatory T cell; nTreg: naturally occurring regulatory T cell.

1.1.1 The Th1 versus Th2 paradigm

The differentiation of naïve T cells into specialised effector helper cells takes place in the peripheral lymphatic system after exiting the thymus (Figure 1.1). This differentiation requires stimulation of the T cell receptor (TCR) by presentation of cognate antigen in the context of the major histocompatibility complex (MHC) II [9]. The direction of polarisation is determined by the cytokine environment of the T cell, which results from the earlier response of the innate immune system to the pathogen present and reflects the signature cytokines it later produces. The presence of an intracellular pathogen leads to production of IL-12 by antigen presenting macrophages and dendritic cells (DCs) [10], which induces Th1 differentiation [11]. Subsequent production of IFNγ from Th1 cells also has an autocrine effect in promoting Th1 differentiation [12]. The presence of an extracellular pathogen leads to production of IL-4, possibly from basophils [13], and the induction of Th2 differentiation. In addition, the T cell growth factor, IL-2, supports the proliferation all T cells [14] and the use of these cytokines has been harnessed for *in vitro* differentiation of Th1 and Th2 cells. The Th1/2 paradigm has been further extended with the discovery of peripherally induced Th17 cells that are characterised by their production of IL-17 [15] and differentiate in response to fungal infection in the presence of TGF-β and IL-6 [16]. An immune suppressive subset, induced regulatory T cells (iTreg), differentiate in the presence of IL-2, upon which they are heavily dependent, and TGF-β [17]. These cells also have a physiologically occurring counterpart termed natural regulatory T cells (nTegs), which exit the thymus as a distinct cell lineage in parallel with naïve T cells [18]. Pathologies are associated with the dysregulation

of each of these subsets, with allergies associated with increased Th2 cytokine production and a Type II response [19] and autoimmune syndromes being associated with increased Th1 [20] and Th17 [21, 22] activity and decreased Treg function [23-26].

1.1.2 Regulatory T cells

The inflammatory response, while often effective at fighting an invading pathogen, is also harmful to host tissues through the production of inflammatory mediators that may enzymatically and chemically degrade tissues. In healthy individuals, these processes are controlled through mechanisms of regulation and tolerance. These mechanisms are required in the resolution phase, when an infection has been cleared and inflammatory processes need to be dampened down, and also to limit immune responses to non-pathogenic foreign antigens (such as food antigens) and self-antigens. Mechanisms of tolerance are categorised as central (occurring in the central lymphatic system) or peripheral (occurring after immune cells reach the peripheral lymphatic system). Central tolerance takes place in the primary lymphoid organs (thymus and bone marrow) during immune cell development. For example, in the thymus, self-reactive T lymphocytes are recognised by the strong affinity of their T cell receptor (TCR) complex for self-antigens presented in the MHC context. These self-reactive T cells are either deleted through apoptosis or are stimulated to replace their TCR by receptor editing [27]. A proportion of self-reactive T cells escape the thymus into the periphery. Here, their activity must be restrained by peripheral tolerance mechanisms. Peripheral tolerance mechanisms include the induction of anergy, which occurs when a self-reactive T cell meets its cognate self antigen in the absence of co-

stimulation. Co-stimulatory signals are produced by ligation of CD28 on the T cell with CD80 and CD86 on an APC. These APC surface molecules are upregulated when the immune system is activated, for example in the presence of an infection or acute inflammation. When inflammatory signals are not present and this co-stimulation is suboptimal, the T cell becomes functionally inactivated.

In an inflammatory environment, when co-stimulation is provided, other mechanisms exist to damped down the immune response. This includes the activity of cells with a specialised function for suppressing immune responses. This is termed ‘dominant’ tolerance and has been demonstrated by the adoptive transfer of T cells [28]. Of these suppressor cells, regulatory T cells are the archetype.

The concept of a specialised cell subset capable of dominant peripheral tolerance arose in the 1970s [29, 30] and was originally highly contentious [31]. Then in 1985, Sakaguchi and colleagues demonstrated that depletion of splenocytes expressing the T cell marker, CD5, resulted in multiorgan autoimmunity, implying that a T cell subset must be actively involved in preventing activation of the immune system against self-antigens [32]. Publications in the 1990s and since have cemented a role for a subset of T lymphocytes with a suppressive specialty [33, 34]. This suppressive subset was further defined in mice by the presence of the T cell co-receptor CD4 and the IL-2 receptor alpha subunit, CD25 [33]. Later work in humans also demonstrated that the CD4+CD25+ T lymphocyte compartment confers dominant peripheral tolerance [35, 36].

In 2003, the molecular control of Treg identity began to be revealed. Work simultaneously published by three independent laboratories demonstrated this regulatory function to be dependent on the forkhead transcription factor, FOXP3 [37-39]. The cell surface phenotype of regulatory T cells (Tregs) has since been further defined with absence of the IL-7 receptor, CD127, enriching for FOXP3-expressors among CD4+CD25+ T cells [40, 41].

1.1.2.1 Suppressive mechanisms of Treg

Tregs limit the activity of a variety of immune effector cells. This includes $CD4^+$ and $CD8^+$ effector T cells, natural killer (NK) cells, natural killer T cells (NKT) [42, 43], antigen presenting cells (APCs) such as dendritic cells (DCs) B cells and monocytes [44], macrophages, mast cells and osteoblasts [45]. Although there is evidence that antigen-specific Treg are more potent [46], Treg have been shown to suppress independently of antigen specificity [47] and through the phenomenon of 'linked suppression', may be activated by their cognate antigen before suppressing responses to other antigens [48]. In addition, 'infectious tolerance' describes the phenomenon by which activated Treg promote suppressive activity in other cells.

There is not a single unified mechanism for Treg suppressive function but instead Tregs appear to exert their suppressive effect in various ways, mediated by both cell contact and soluble factors. Some of these are illustrated in Figure 1.2 and are described in the following paragraphs. I have used the term T responder (Tresp) throughout this thesis to

distinguish Treg from other effector T cells with possible pro-inflammatory activities.

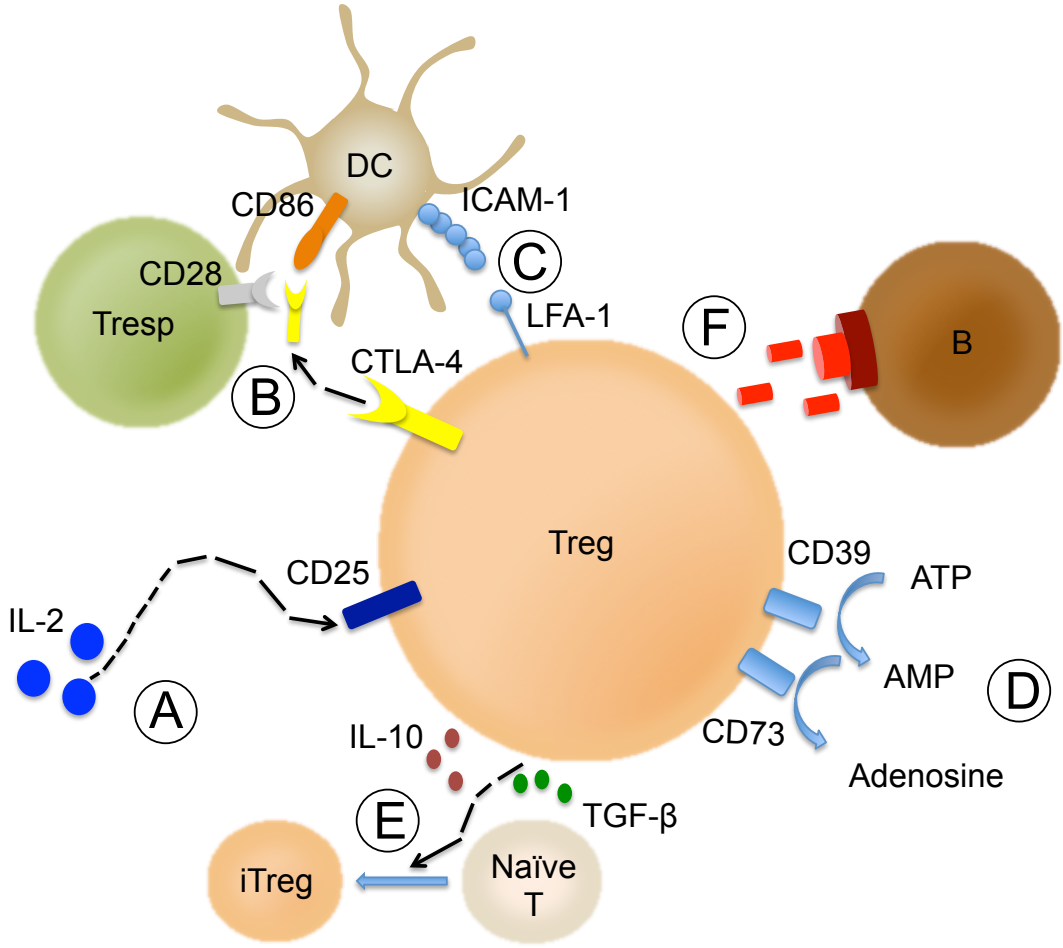


Figure 1.2: Suppressive mechanisms of CD4 FOXP3+ Treg.
The suppressive mechanisms employed by Tregs are diverse and affect several effector cells. A: Surface expression of cytokine receptors, such as the IL-2 receptor alpha subunit, CD25, sequesters pro-inflammatory molecules. B: CTLA-4 in its surface bound and soluble form competes with CD28 on Tresp to bind the co-stimulatory molecules CD80 and CD86 from antigen presenting cells (APCs) such as DCs. Ligation with these receptors also results in their down regulation and the promotion of apoptosis in APCs. C: Adhesion molecules such as lymphocyte function associated antigen 1 (LFA-1) promote the aggregation of Tregs around APCs and their spatial competition with Tresp. D: CD39 and CD73 catalyse the breakdown of pro-inflammatory ATP to anti-inflammatory adenosine. E: Secretion of cytokines such as TGF-β and IL-10 provide 'infectious tolerance' by promoting suppressive function in other cells, for example through the induction of naïve cells into induced Treg. F: The cytolytic pore-forming enzymes perforin and granzyme allow Treg to induce apoptosis in target cells. Tresp: T responder; DCs: dendritic cell; iTreg: induced Treg.

1.1.2.1.1 Cell-contact mediated mechanisms

Cytotoxic T-Lymphocytes Antigen-4 (CTLA-4) is a T cell surface molecule analogous to the co-stimulatory molecule CD28 in that it ligates with CD80 and CD86 on APCs. However, in contrast to CD28, CTLA-4 ligation results in the transduction of inhibitory rather than stimulatory signals. This leads to reduced activation of the T cell itself but its expression on Treg has also been shown to have a role in Treg suppression of other cells. Our lab has shown that reduced surface CTLA-4 on Treg leads to reduced suppression of IFN-γ production by T responders [49]. In addition, competition of Treg with Tresp for ligation of CD80 and CD86 on APCs leads to a reduced likelihood of Tresp activation. Similarly, the Treg-APC interaction may result in capture of CD80 and CD86 and a reduced ability for APCs to further interact with other T cells [50]. This interaction may also have APC-intrinsic effects with signal transduction downstream of CD80 and CD86 leading to downregulation of these molecules and even apoptosis [51].

Extracellular adenosine metabolism by the Treg derived CD39 (ectonucleoside triphosphate diphosphohydrolase 1, ENTPD1) and CD73 (ecto-5'-nucleotidase) also reduces Tresp activity. These cell surface molecules catalyse the conversion of ATP to AMP and AMP to adenosine, respectively. The resulting adenosine acts as an anti-inflammatory mediator, signalling through adenosine receptors located on the surface of Tresp [52, 53]. Ligation with the adenosine receptor results in increased intracellular cAMP in Tresp and a concomitant reduction in their function [54]. This pericellular production of adenosine by Tregs has been shown to suppress the proliferation of activated T effector cells [55].

By promoting the aggregation of Treg around APCs, adhesion molecules may also be important in reducing inflammation by spatial competition with Tresp for interaction with APCs [51]. For example, lymphocyte function associated antigen 1 (LFA-1) is an integrin found on several immune cell lineages and is involved in recruitment to sites of inflammation through affinity to ICAM-1. Treg lacking LFA-1 have been shown to have a reduced function *in vivo* [56].

Other extracellular Treg molecules with roles in suppression include the lymphocyte activation gene 3 (LAG-3) and glucocorticoid induced tumour necrosis factor receptor (GITR), which are both increased upon activation. The mechanistic role of LAG-3 in Treg function is not known but expression is increased upon culture with Tresp and blocking of LAG-3 inhibits Treg suppression while ectopic expression confers suppressive activity [57]. GITR is found at a high level on Treg. However, elucidation of its role and function is complicated by the fact that it is expressed on other immune system components and is upregulated on Tresp upon activation. While its role in the suppressive function of Treg is unknown, signalling via GITR has been shown to expand the Treg population while retaining function [58].

1.1.2.1.2 Soluble factor mediated mechanisms

The inhibitory effects of CTLA-4 described above may also be produced by secretion of soluble CTLA-4 from Treg [51]. Other soluble factors include perforin and granzyme, components of cytoplasmic granules also used by cytotoxic T cells and NK cells. Perforin triggers result in pore formation in

target cells, allowing the induction of apoptosis through caspase activation as granzyme is released into the cell through these pores [59-61].

Being defined by high expression of the IL-2 receptor alpha subunit, CD25, Tregs are thought to limit inflammation by sequestering this T cell growth factor away from Tresp, thus limiting their growth [60]. This same mechanism of sequestration may also apply to other pro-inflammatory cytokines such as IL-1 and TNF-α [62, 63].

Tregs also contribute to the cytokine milieu through the secretion of anti-inflammatory cytokines including transforming growth factor-beta (TGF-β), IL-10 and IL-35. TGF-β influences activation, proliferation and differentiation of target cells, for example it has been shown to inhibit Th1 differentiation and to suppress CD8+ cytotoxic T lymphocytes [64, 65] and NK cells [43]. IL-10 is predominantly secreted from non-FOXP3 expressing suppressor cells i.e. Tr1 cells [66]. However, increased production of IL-10 by FOXP3+ Tregs has also been thought to protect against the mouse model of multiple sclerosis, experimental autoimmune encephalomyelitis (EAE) [67-69]. Treg IL-10 production may have effects on APC and NK function [70]. IL-35 production from FOXP3+ Tregs has been shown to suppress Tresp proliferation and to limit mucosal inflammation *in vivo* [68, 71, 72].

In addition to inhibiting the action of immune effectors, these anti-inflammatory cytokines also promote suppressive function in other cells and so amplify their suppressive effects through infectious tolerance. This

includes through the differentiation of naive T cells into FOXP3 expressing and suppressive iTregs in the presence of TGF-β [63, 73]. In addition, TGF-β has been shown to skew Tresp to a suppressive Tr1 phenotype [74].

This collection of mechanisms employed by Treg may reflect the presence of different physiological subsets of Tregs, some of which are discussed in more detail below.

1.1.2.2 Subsets of CD4+FOXP3+ Tregs

The expression of molecules associated with Treg function varies under different activation conditions or environmental and physiological contexts. In addition there are no cell surface markers that are limited to Treg only. Even CD25 and FOXP3 are increased in Tresp upon activation, which can be problematic for identifying and isolating Treg. This lack of unifying features hints at heterogeneity in the CD4+FOXP3+ population.

1.1.2.2.1 Naturally occurring and peripherally induced Tregs

'Naturally occurring' Tregs (nTregs) represent a specialised subset of T cells that is developmentally determined in the thymus before activation in the periphery through the encounter of antigen [75]. These nTreg exit the thymus expressing FOXP3 alongside undifferentiated naïve CD4 T cells. In the appropriate cytokine context, IL-2 and TGF-β, activation of naïve CD4 T cells induces expression of FOXP3 and suppressive activity as these cells differentiate into iTreg [76-79]. In addition, human *in vivo* evidence shows some regulatory CD4+FOXP3+ cells may arise from CD4+FOXP3-

memory, as opposed to naïve, cells [80]. nTregs and iTregs may have separate and synergistic roles *in vivo* [81, 82]. iTreg differentiation may be particularly relevant at sites of inflammation where they are required to curtail the potentially harmful effects of an immune response [68, 83]. This includes in gut-associated lymphoid tissue (GALT), where induction of tolerance is necessary in the presence of foreign materials such as commensal bacteria and food antigens [84]. There is mechanistic support for the *in vivo* generation of iTreg from studies in mouse [17, 76, 85]. *In vitro* generation of iTreg has therapeutic possibilities through the induction of Treg from naïve T cells with the subsequent return to the body. However, while *in vitro* culture has lead to successful Treg induction from mouse cells, FOXP3 expression in human cells in this context is not synonymous with suppressive ability. These FOXP3-expressors even secrete the T cell growth factor, IL-2, implying they are pro-inflammatory, and it has been suggested that FOXP3 in this context may reflect general T cell activation rather than a lineage with suppressive function [86]. This implies there may be another signal required for human iTreg generation. Several groups have shown that alternative protocols involving all-trans retinoic acid (atRA) [87] and stimulation of the aryl hydrocarbon receptor (AhR) through dioxin-like ligands such as TCDD (2, 3, 7, 8-Tetrachlorodibenzodioxin) [88] produce functionally suppressive Treg. The differences between Treg of mice and humans may reflect different requirements to support stability and reduce plasticity of Tregs and might also suggest the phenotype orchestrated by FOXP3 may differ between the two species.

The transcription factor Helios has been suggested as a marker to distinguish nTreg from iTreg, with its expression being restricted to the former [89]. As both mouse and human *in vitro* induced Treg lack Helios expression, it was suggested that Helios is not affected by TCR stimulation and therefore is not a marker of activation, in the way that FOXP3 is [89]. However, since this assertion was made, other evidence has shown Helios expression to reflect the context of stimulation [90, 91]. In addition, Helios has been associated with Tresp lineages [92]. Work in T cell lines has shown Helios may have a potential role in Treg function through binding of the FOXP3 promoter, as identified by ChIP-Seq, with knockdown of Helios leading to downregulation of FOXP3 expression [93].

1.1.2.2.2 Naïve and memory Tregs

CD4 T cells exit the thymus in a naive state being unactivated and not terminally differentiated. When the TCR of a naive T cell ligates with its cognate antigen in the context of MHCII in the peripheral lymphoid organs, it becomes activated and proliferates. This results in clonal expansion of effector T cells with specificity to the activating antigen. These antigen-stimulated lymphocytes then patrol the periphery to enact this response where it is required when re-encountering the same antigen. After clearing an infection and inflammation is resolved, the number of T cells specific to the antigen decreases until only a residual population remains. These memory cells reside in the bone marrow until promoted to expand and respond once again upon recognition of the same epitope, allowing a faster response to the secondary infection therefore conferring immunity.

Naïve CD4 T cells exiting the thymus express the cell surface protein

tyrosine phosphatase receptor CD45RA. This is the largest isoform of the gene PTPRC, the alternative splicing of which is linked to activation of naive T cells. When the TCR is stimulated by ligation with antigen, the T cell undergoes alternative splicing of the PTPRC gene, with differential expression of isoforms delineating maturational status. Eventual maturation to a memory phenotype is denoted by expression of the RO isoform. Therefore, these isoforms can be used to distinguish between naïve and memory CD4 T cells.

Within the Treg population, expression of CD45RA has demonstrated phenotypic and functional heterogeneity within the Treg population. For example, CD45RA-negative Tregs show decreased CpG methylation at the *RORC* locus, which encodes the Th17 lineage specifying transcription factor, RORγ, implying more plasticity at the level of transcriptional regulation in memory compared to naïve Treg [94, 95]. Miyara and colleagues [96] performed a phenotypic comparison of Treg varying in CD45RA and FOXP3 expression, delineating several subsets. They concluded that FoxP3^{lo}CD45RA+ cells represent resting Tregs. Upon *in vivo* activation these expand and become FoxP3^{hi}CD45RA-. Both of these populations are potently suppressive but this expansion appears to result in rapid death of these Tregs after exerting their suppressive function. In addition FoxP3^{lo}CD45RA- cells lack suppressive function, express RORC and secrete the pro-inflammatory cytokines IL-2 and IFNγ, so these may represent recently activated effector T cells transiently expressing FOXP3.

1.1.3 The relevance of Treg lineage dysregulation for autoimmunity

Treg are necessary for peripheral tolerance of the immune system to self-antigens. Mutations disrupting the Treg-specifying transcription factor FOXP3, occur in the IPEX (immunodysregulation polyendocrinopathy enteropathy X-linked) syndrome and its murine counterpart the Scurfy phenotype, lead to an absolute lack of Treg and systemic autoimmunity [25]. Ablation of Treg in *in vivo* models leads to autoimmune pathology involving multiple organ systems including gastritis, thyroiditis, diabetes and inflammatory bowel disease [23, 25], and is fatal [26]. Disease is ameliorated by adoptive transfer of Tregs to Treg-depleted mice [24] and to mouse models of autoimmune diseases, such as non-obese diabetic (NOD) mice (a type I diabetes model) [97] and experimental autoimmune encephalomyelitis (EAE, a multiple sclerosis [MS] model) [98]. Co-culture experiments have demonstrated a reduced ability of Tregs from MS patients to suppress the proliferation of T effectors from healthy individuals [99]. Tregs from individuals with T1D have also demonstrated a reduced responsiveness to IL-2, which correlates with loss of FOXP3 expression [100]. Post-mortem histological analysis has shown that FOXP3+ Tregs are rare in the pancreatic islets of patients with type 1 diabetes (T1D) implying that inadequate numbers of Tregs at the site of inflammation are in part responsible for the disease [101]. In MS, Treg numbers appear to be increased in both cerebrospinal fluid [102] and peripheral blood [102-104] and therapy normalises this increase [105]. This could indicate that in the context of autoimmunity, not all FOXP3-expressing CD4 cells are suppressive [106].

1.1.3.1 The clinical presentation of systemic lupus erythematosus (SLE)

Systemic Lupus Erythematosus (SLE) is the archetypal systemic autoimmune disease. It is distinguished from other systemic autoimmune diseases by the presence of auto-antibodies to nuclear components in over 95% of patients [107]. These nuclear components include double-stranded DNA and nucleosomes but also antigens with unknown biological roles, such as 'Ro' and 'Sm', the nomenclature of which has been based on the lupus patients in which they were first described [107].

Patients present with diverse clinical manifestations and so it is has been debated whether SLE should be considered a single disease or a collection of diseases. Auto-antibody repertoire reflects clinical presentation with reactivity being relevant to the organs and tissues involved in pathology [108]. The diversity of organ involvement and tissue destruction, as illustrated in Figure 1.3, can include arthritis, nephritis, dermatological lesions, haematological abnormalities (eg. anaemia and thrombocytopenia), cardiovascular problems and neurological complications (ranging from seizures to psychosis). Chronic and extreme fatigue is also a common and debilitating symptom. In addition, the classical physical defining feature of lupus is the 'butterfly rash' across the face, which was considered to resemble the bite of a wolf when it heals [109]. It was this distinctive mark that lead to recognition of lupus in the 18th Century and gave the disease its name from the Latin for wolf [109].

Patients also vary with respect to disease severity, control of disease

through treatment and in their pattern of disease flares. Disease activity is measured by anti-DNA antibody titre, a value of 50 IU (International Units) being considered normal and an increase seen with disease activity, and complement C3 level, 0.9 g/L being considered normal and a decrease being seen with disease activity. In addition, classification systems have been designed to aid standardisation of descriptions of disease severity. This includes the British Isles Lupus Assessment Group (BILAG) index, through which disease progression is assessed by comparison of clinical features and organ involvement with previous observations. Scores are based on whether the clinical feature was present in the previous visit or is new and whether the feature has improved or worsened. The Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) is a similar system but clinical features are scored without reference to clinical history.

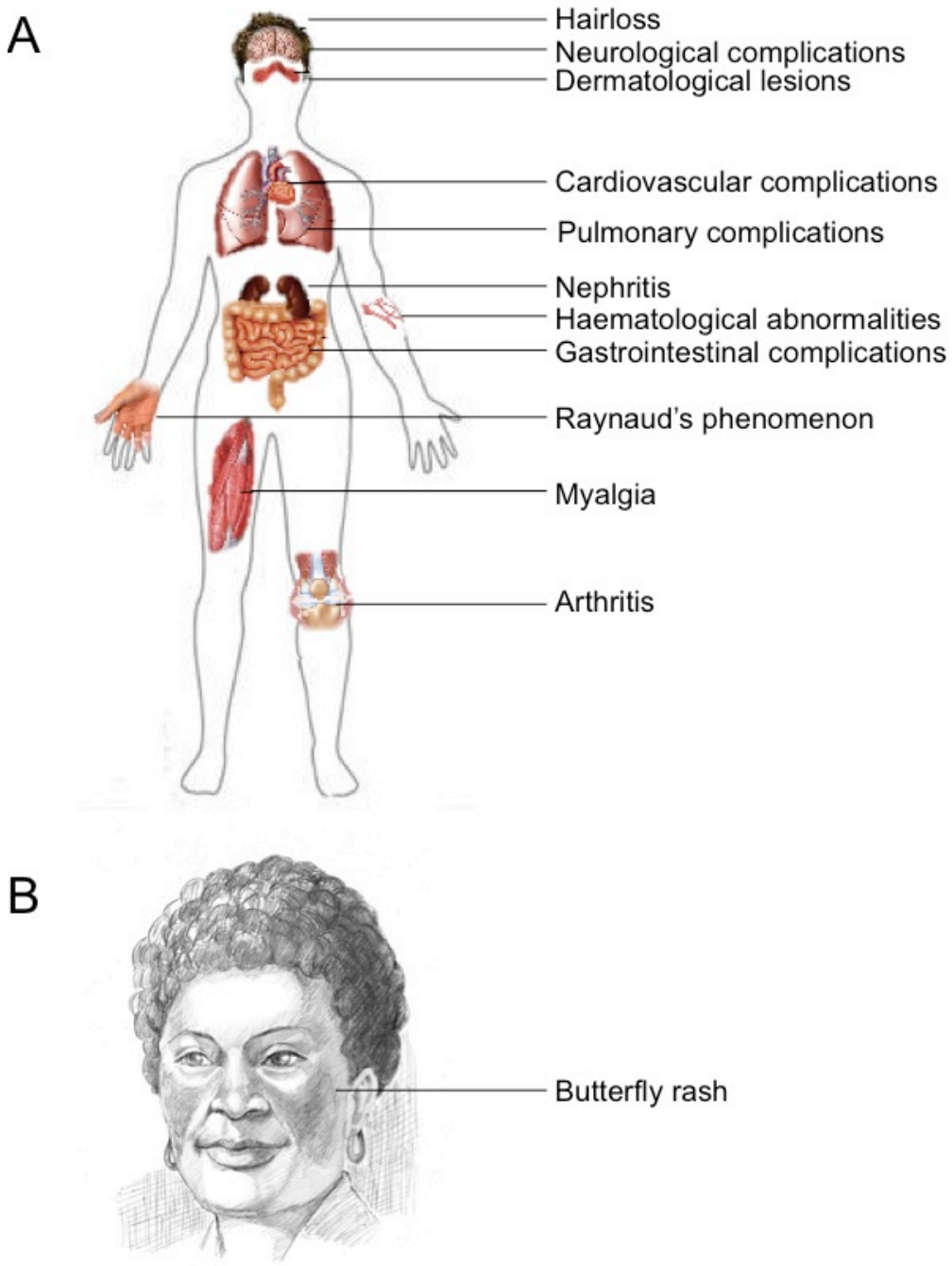


Figure 1.3: The Diverse Clinical Presentations of Systemic Lupus Erythematosus
SLE patients exhibit heterogeneous clinical presentations. A: Organ involvement and tissue destruction can result in arthritis, nephritis, dermatological lesions, haematological abnormalities, cardiovascular problems and neurological complications. B: The classical 'butterfly rash', which was considered to resemble the bite of a wolf as it heals. Images adapted from the image gallery of National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) [110].

1.1.3.2 The epidemiology of SLE

As is common for autoimmune diseases, there is no cure for SLE and its cause in not fully understood. Reported prevalence is 20 to 150 cases per 100,000 people (0.02-0.15%) [111], however SLE is 9 times more common in women than men and the incidence rate amongst women is 164 (Caucasian) to 406 (African American) per 100,000 women (0.16-0.41%) [111]. Onset of disease typically occurs between 16 and 55 years of age (65% of patients) [111], but is not limited to this age bracket [111, 112]. Survival rate in the 1950s was estimated to be 4 years for 50% of patients [108]. Worldwide, death as a direct result of disease activity (for example through organ failure) has declined [113] but mortality continues to arise from complications. These complications include infections associated with non-specific immunosuppressive treatments [114], increased cardiovascular risks [113, 114] and cancer (often non-Hodgkin's lymphoma and lung cancer) [113, 114]. The survival rate has improved to 80% at 15 years. However, this more favourable statistic still however implies that a lupus patient diagnosed at the age of 20 has a 1 in 6 chance of dying by the age of 35 [108]. These statistics also highlight the longevity of this disease that requires decades of medical intervention.

The gender difference indicates potential aetiological roles for hormonal and genetic factors. Hormonal involvement is supported by the common age of onset being during reproductive years and reported associations with oral contraceptives, post-menopausal hormone replacement therapy, pregnancy and breastfeeding [111]. X chromosome involvement is suggested by the presence of SLE associated variants (including in the

genes *IRAK1, MECP2* and *TLR7*) [111], an increased prevalence in men with Klinefelter's syndrome (XXY) and a decreased prevalence in women with Turner's syndrome (XO) [111]. Clearly gender is not the only influence as not all women are affected and men are not resistant. However, there are interesting parallels between the higher prevalence of autoimmunity in women in general [115], the female specific need for foetal tolerance, an evolutionary association of peripheral Treg induction with placental mammals [116] and the encoding of the Treg lineage-specifying transcription factor FOXP3 on the X chromosome.

Further support for the role of genetics is supported by ethnic and familial associations [111], including a high concordance rate (14 to 57%) of monozygotic twins [117, 118]. Risk alleles exist in immunologically relevant genes, including those for complement and innate immunity components, lymphocyte signalling proteins and the major histocompatibility locus (MHC). However, genetic associations account for only 18 percent of SLE incidence [111] and so environmental and epigenetic factors also contribute.

Environmental associations have included factors thought to reflect immune system disturbance, such as pollutants, including silica, and allergies to medications [111]. There are links with exposure to ultraviolet (UV) light and this has particular relevance for patients with dermatological involvement [111]. Viral infections have also been associated with SLE. This includes higher antibody titres and viral loads relating to Epstein-Barr virus (EBV) [111] and also suggestions that viral infection or mimicry by

endogenous retroviruses may represent triggering events. Interestingly, a viral response signature, including interferon production is associated not only with SLE but with other fatigue related syndromes [119-121].

1.1.3.3 Treatment of SLE

Standard treatment for SLE begins with non-specific anti-inflammatory agents including disease-modifying anti-rheumatic drugs (DMARDS), such as hydroxychloroquine, azathioprine, leflunomide, methotrexate, mycophenolate mofetil, cyclophosphamide and cyclosporine [122]. This is accompanied by administration of non-steroidal anti-inflammatory drugs (NSAIDs) and corticosteroids [122]. The broad action of these therapies results in long-term toxicity and immune-compromisation that manifest as side effects including osteoporosis and increased risk of opportunistic infection [122]. The use of these non-specific therapies is far from ideal and is limited if possible [122-124].

Patients with greater disease activity and those not responsive to the first line therapies described above may progress to treatments with greater associated costs. This includes monoclonal antibody therapies aimed at depleting abnormal or overactive immune components such as rituximab and belimumab, which have been licensed for use in other clinical contexts and are directed at limiting the activity of B cells [125]. Rituximab, which is directed against the mature B cell marker CD20 and results in B cell depletion, has been approved for use in other autoimmune diseases and also shows clinical success in SLE [124, 126, 127]. However, two randomised controlled trials have not been successful [128, 129]. Belimumab is a monoclonal antibody against B-cell activating factor

(BAFF) and leads to inhibition of B-cell survival and differentiation. This results in reduction of autoantibodies, which correlates with reduced disease activity in SLE [130]. In March 2011, the Food and Drug Administration (FDA) approved the use of anti-BAFF monoclonal antibody, belimumab, for treatment of SLE in the United States [124] and the European Medicines Agency (EMA) has recommended the granting of marketing authorisation [131]. In the United Kingdom however, the National Institute for Health and Clinical Excellence (NICE) has rejected the use of belimumab as the perceived health benefits are outweighed by costs [122].

Hematopoietic stem cell transplantation (HSCT) has been used for re-education of the immune system in SLE. Transplant-mediated mortality is however a concern and so HSCT remains a possibility only in cases of aggressive disease where other possibilities are exhausted, rather than as a routine treatment [132, 133].

1.1.3.4 The immune pathology of SLE

The pathogenesis of SLE is multifactorial but culminates in the loss of tolerance to self-antigens, followed by activation and expansion of auto-reactive lymphocytes with the concomitant production of pro-inflammatory mediators and autoantibodies. This eventually leads to systemic inflammation and tissue damage.

The autoantibody repertoire of SLE patients is diverse as it reflects heterogeneous organ involvement but reactivity against nuclear antigens is a defining feature of SLE that is shared among 95% of patients. This

characterisation of SLE on the basis of autoantibody production has strongly implicated a role for B cells in SLE pathogenesis. In support of the central role of B cells in SLE, patients show differences in B cell maturation and class switching, in addition to repertoire [134]. These B cell abnormalities may potentially be secondary to inflammation rather than intrinsic defects but, in spite of this, B cell targeting therapies, such as rituximab and Belimumab (described in Section 1.1.3.3), are attractive options.

A theory for the abundance of anti-nuclear antigen antibodies is that inefficient clearance of apoptotic cellular debris leads to exposure of nuclear epitopes and the production of immune complexes against these. In support of this, increased apoptosis is seen in lymphocytes from SLE patients, possibly due to their increased activation [135] and accumulation of post-apoptotic debris is seen in B cell germinal centres of lymph nodes of SLE patients [136]. Classical complement components (C1q, C2, C4) aid the clearance of immune complexes and autoantigens and, consistent with this, deficiencies in these early complement components predispose towards SLE [137]. Activity of later components of the complement cascade, however, contributes towards pathogenesis with hydrolysis of alternative complement pathway component, C3, leading to downstream promotion of inflammation and tissue injury [137]. The serum level of C3 is therefore associated with immune activation and is indicative of active disease in SLE patients.

In addition to complement, further dysregulation of the innate immune

response is seen in SLE, with increased signalling downstream of the Toll-like receptors (TLRs) 4 [138], 7 [139] and 9 [140] associated with the initiation and perpetuation of SLE [141]. These TLRs recognise lipopolysaccharide, single-stranded RNA and unmethylated CpG sequences, respectively, the latter two again reflecting the involvement of nuclear epitopes. The activity of anti-viral-response associated receptors is echoed in the cytokine signature seen in SLE patients.

SLE patients demonstrate a unique inflammatory cytokine environment, particularly regarding high levels of the anti-viral response associated cytokine, interferon-alpha (IFN-α) [142], and IFN-α inducible genes [143-146]. The primary producer of IFN-α appears to be plasmacytoid dendritic cells (pDCs) [137, 147]. This IFN-α production drives autoimmunity by promoting the maturation and pro-inflammatory role of DCs [148], the differentiation of plasma cells [149] and also B and T cell signalling [148]. The serum from SLE patients is also high in the pro-inflammatory cytokines IL-6 (from monocytes, fibroblasts and endothelial cells in addition to lymphocytes) [150, 151] and IL-17 (from neutrophils, NK cells, CD4, CD8, double negative and γδ T cells) [151, 152] and also the pleiotropic cytokine TNF-alpha [150], the role of which is poorly understood although levels correlate with disease activity [153].

The Th17 CD4 T helper cell lineage has been implicated in SLE pathogenesis through the increased production of the pro-inflammatory cytokine IL-17 [154] and reduced IL-2, which inhibits Th17 differentiation [155]. In addition, the role of T-lymphocyte help for autoantibody

production by B cells is critical to the pathogenesis of SLE. This includes in class-switching and costimulation by T follicular helper cells [156]. In addition, signalling through the T cell immunological synapse is aided by an abnormal membrane lipid aggregation and profile in SLE [157].

1.1.3.5 CD4 T cell transcriptional abnormalities in SLE

Gene expression profiling of PBMCs and whole CD4 T cells reflects the IFN-α signature that is characteristic of SLE [158, 159]. Acetylation level of global H3 and H4 is reduced in SLE CD4 T cells, as is the global H3K9 methylation level, while global H3K4 methylation is not different to healthy individuals [160]. In correlation with this, mRNA levels of the histone acetyltransferases CBP and P300, the histone deacetylases HDAC2 and HDAC7, and histone methyltransferases SUV39H2 and EZH2 are decreased [160]. However, activity of these histone modifiers and the specific genomic sites that exhibit differences in post-translational histone modifications are unknown [161]. Increased H3K27me3 at the promoter of hematopoietic progenitor kinase 1 (*HPK1*) has been observed in CD4 T cells, which correlated with an expected decrease in mRNA and protein. [162]. In terms of T cell relevant transcription factors, STAT1 has been implicated in SLE pathogenesis by its role in interferon signalling, however total and phosphorylated levels do not correlate with SLE [163, 164]. In whole T cells (pan-CD3), phosphorylated STAT5 levels are increased above healthy controls and rheumatoid arthritis patients [163].

1.1.3.6 Treg in SLE pathogenesis

As described above, the immune pathogenesis of SLE is multifactorial and is underpinned by a general dysregulation of inflammatory mediators. As

Treg are essential for limiting deleterious autoimmune reactions in the periphery, they have been investigated for their relevance to SLE pathogenesis. This has uncovered some evidence of both quantitative and functional differences in the Treg of SLE patients compared to healthy individuals and other disease controls.

1.1.3.6.1 Quantitative differences

Accounts of numerical differences in Treg vary according to phenotypic description [165]. When identifying Treg using the basic criteria of CD4+ and CD25+, several studies find that Treg are depleted in SLE patients [166, 167], both with [168] and without [169] correlation with disease severity. Another study showed a non-significant trend for increase of CD4+CD25+ T cells, with numbers significantly correlating with disease severity but also with corticosteroid treatment, suggesting that evaluating numbers of Treg is confounded by treatment rather than directly related to disease [170]. When Treg are defined as CD25 high or 'bright', the majority of studies see that they are decreased in patients [171-176] but a minority still observe that numbers compare with healthy individuals [177, 178] or even that they are increased [179]. More recent studies that exclude CD25+ cells expressing CD127, a population that is enriched for FOXP3+ 'true' Treg [40], find a decrease in Treg that correlates with increased numbers of Th17 cells [180, 181]. When FOXP3 expression is used to define Treg, regardless of CD25 or CD127 expression, the majority of studies, including data from our lab [182], show an increase [174, 183, 184], with some still showing normal numbers [167, 185]. When FOXP3 is assessed in the CD25+ compartment, increased numbers are seen [186]. However, when analysis is confined to the CD25 high population and

numbers of FOXP3 expressing cells examined, the converse is found and decreased numbers are observed [187, 188], but again there are exceptions [189]. CD127- Treg and CD25 high FOXP3+ Treg, which are likely to represent the same population, show a general consensus of a decrease in numbers in SLE. The comparisons with CD25- Treg indicate that in SLE patients FOXP3 expression is more frequent in CD25- CD4 T cells compared to a healthy setting, while in CD25+ cells it is decreased. This suggests instability to the Treg phenotype. As Treg are highly dependent upon IL-2 for survival and FOXP3 expression [190, 191], the impaired production of IL-2 seen in SLE patients [192] supports the possibility of Treg defects that relate to the expression of the IL-2 receptor, CD25.

Decrease in numbers of Treg may be due to deficient production, maintenance, stability or survival of Treg and potentially may arise prior to disease onset but also in response to the inflammatory environment of the disease itself or to treatment. Increases in Treg numbers may reflect the expected physiological response of an attempt to limit reactivity against self-antigens. This could be interpreted as Treg being functionally equivalent to those from healthy individuals, but with the extent of inflammation, and/or the presence of defects outside of the Treg compartment, being too extreme for Treg to restrain. Alternatively, increased numbers may reflect Treg expansion in response to inflammation, even though the cells exhibit defective suppressive capacity.

1.1.3.6.2 Functional differences

As with investigations into numerical differences, there is conflicting

evidence on functional defects in Treg in SLE. As assessed by the *in vitro* suppression assay, the CD25-high population has been shown by some to be normal in terms of suppression of Tresp proliferation [172, 185, 188] and cytokine production [172], with some evidence showing a defect resides within the Tresp rather than Treg population [185]. However, defects in this population have also been observed in the suppression of proliferation [176], but not in all patients [177] and these defects correlate with disease activity [193]. Defective suppression is also seen in the total CD25 population regardless of therapy [166]. In addition, functional Treg from both healthy individuals and SLE patients have been shown to be rendered defective following culture with APCs from SLE patients [186].

Identifying functional defects in Tregs in SLE is complicated by their low frequency in peripheral blood (1% of PBMCs), which is decreased further under the lymphopenic condition of active disease. This is compounded by the heterogeneous nature of SLE, because any defects found may not be applicable to all patients, and the likelihood that examining peripheral blood Tregs will not fully reflect those situated in inflamed sites. In addition, the heterogeneity of Treg themselves, and the possibility of Treg subsets with differences in phenotype and suppressive mechanisms, further complicates functional assessment. There is no accepted Treg-specific marker that can be used for identification or isolation; FOXP3 and CD25 are also expressed, at least temporarily, in activated responder T cells without conferring suppressive function [194].

In support of defective Treg activity in SLE, the inflammatory cytokine

environment has been shown to impact on Treg function. For example, IFN-a production from APCs rendered Treg dysfunctional [186]. Serum from SLE patients is characterised by high levels of IL-6 [150] and TNF-a [150, 153], both of which may interfere with Treg function; IL-6 promotes the conversion of Tregs to an inflammatory Th17 phenotype [195], which has been implicated in other autoimmune diseases [196]. Correspondingly, levels of the Th17 cytokine IL-17 have also been found to be increased in SLE patients [152]; TNF-alpha, through TNF receptor II mediated signalling, can down-regulate Foxp3 expression in Treg and hence potentially diminish suppressive capacity [197]. Additional support for potential Treg defects comes from a recent SNP association study between sub-phenotypes of the disease and overactive variants of known FOXP3 target genes [198].

1.1.4 Development and differentiation of T cells

Below I describe the intracellular events downstream of developmental and environmental stimuli that facilitate T cell differentiation. These signalling events lead to the activation of transcription factors that have common roles among all T cell subsets but also those that promote lineage specification.

1.1.4.1 TCR signalling

Signalling events downstream of antigen presentation lead to activation of the T cell through several pathways. Engagement of the CD4 co-receptor with MHCII from an APC leads to lymphocyte-specific tyrosine kinase (LCK) binding to the intracellular domain of CD4 (reviewed in [199]). This brings LCK into close vicinity with the CD3 intracellular chains leading to

phosphorylation of intracellular tyrosine-based activation motifs (ITAMs) enabling the recruitment and activation of Zeta-chain-associated protein kinase 70 ZAP70 [199]. This is followed by phosphorylation of the linker for activation of T cells (LAT) and the formation of a complex of multiple signalling proteins. Events downstream of this LAT 'signalsome' lead to electrochemical signalling, via the release of calcium ions (Ca^{2+}) from the endoplasmic reticulum, and activation of the mitogen-activated protein kinase (MAPK) kinase and nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) signalling pathways [199].

Concomitant signalling through CD28, as well as CD3, by the co-stimulatory molecules CD80 and CD86 is required to avoid T cell anergy. Phosphorylation of intracellular motifs of CD28 recruits phosphatidylinositol 3-kinase (PI3K), which targets AKT and leads to activation of transcription factors including NF-κB, NFAT and mammalian target of rapamycin (mTOR) [200]. CD28 signalling also leads to protein kinase C theta (PKC-θ) phosphorylation followed by activation of c-Jun N-terminal kinases (JNK) and extracellular signal-regulated kinases (ERK) 1 and 2 [200]. There are differential requirements for CD3 and CD28 signalling between Treg and Tresp development. For example, signalling molecules downstream of mTOR, specifically p70S6 kinase (p70S6K) and 4E binding protein 1 (4E-BP1), show reduced phosphorylation in Treg compared to Tresp [201]. This is consistent with the use of rapamycin, which inhibits the mTORC1 subunit of mTOR, as an immunosuppressant in clinical use and also for the preferential expansion of Treg over Tresp *in vitro* [202].

1.1.4.2 Treg specific development

Treg develop in the thymus as a lineage distinct to naïve Tresp and are also induced in the periphery from naïve T cells. In the thymus T cells go through a selection process to eliminate those with a self-reactive TCR, the result of the somatic gene rearrangement that is necessary for the diverse TCR repertoire that allows us to acquire immunity to limitlessly evolving pathogens. T cells with self-reactivity are stimulated to undergo receptor editing or programmed cell death in the process of negative selection. This process is not completely efficient and self-reactive T cells do escape the thymus including in healthy individuals. The presentation of self-antigen by thymic epithelial cells has been proposed to also positively select for, or promote the development of, Treg. There is evidence of TCR reactivity to self-antigen in the Treg population [203, 204] and it has been suggested that this self-reactivity among immuno-suppressive nTreg leads to their activation in the periphery, guarding against autoimmunity. Escape of these self-reactive nTreg from the thymus may be reliant on a lower avidity for self-antigen below the threshold for negative selection [205, 206]. Signalling downstream of this TCR engagement may also promote the development of Treg function in thymocytes [207]. The downstream consequences of TCR signalling are discussed in the following paragraphs. High expression of CD25 suggests Treg are dependent on its ligand, IL-2, and this is shown by the development of spontaneous autoimmunity in IL-2 deficient mice [208]. In the periphery, *de novo* generation of iTreg from naïve T cells also involves TGF- β [17].

Treg are dependent upon IL-2 binding CD25, the alpha subunit of the

receptor, which is highly expressed and is used to define and isolate Treg. This alpha chain of the receptor is not thought to transduce signalling directly but dimerises with the beta (CD122) and gamma (CD132) chains to form the IL-2 receptor complex. The beta and gamma chains are phosphorylated by Janus kinases (JAK) 1 and 3, respectively [209], leading to recruitment and activation of the transcription factor signal transduce and activator of transcription (STAT) 5 [210].

TGF-β signalling promotes the differentiation of Treg over pro-inflammatory lineages. Upon TGF-β binding to TGF-β receptor II (TGF-βRII), receptor dimerization with TGF-βRI leads to recruitment and activation of receptor-regulated SMADs (R-SMADs) 2 and 3 [211, 212]. These have been suggested to complex with common-mediator SMAD (co-SMAD) 4 followed by translocation of this transcription factor tripartite complex to the nucleus [213, 214]. Activation of SMAD independent signalling pathways may also occur downstream of the TGF-β receptors. This is poorly characterised but may lead to activation of JNK, p38 MAPK and NF-κB downstream of MAPK kinase kinases (MAPKKKs) [215, 216]. These events may converge on SMAD pathways to regulate SMAD signalling [216].

Paradoxically, signalling through receptors for the pro-inflammatory cytokine TNF-α promotes Treg activity. The stabilisation of Treg phenotype, function and expansion [217, 218] in response to TNF production by pathogenic Tresp may allow for a negative feedback loop that limits Tresp activity [219]. This has been shown to be mediated

through TNFR2 [217, 219, 220] and other members of the TNF receptor superfamily, including OX40 and 4-1BB [221].

Ultimately, these signalling cascades link the cells external environment to rearrangement of the chromatin landscape and changes in gene expression, allowing adaptation of the T cell phenotype to an environmental stimulus [199].

1.2 Transcriptional control of T cell differentiation

Despite the nearly identical DNA sequence present in the genes of each cell of an organism, distinct cellular identities are formed from the increased activity of cell specific genes and the repression of genes relevant to alternative cell fates. For example, in Th1 cells *IFNG* is upregulated while *IL4* is downregulated and vice versa for Th2 cells. The maintenance of this cell specific gene expression over cell divisions allows the inheritance of cell identity and the establishment of cell lineages. Control of gene expression at the level of transcription provides a mechanism for this epigenetic inheritance and influences cell identity. This level of regulation therefore has consequences for normal cellular processes, such as fighting invading pathogens, and for disease when this differentiation becomes dysregulated, for example in autoimmunity.

1.2.1 Transcription factors controlling T cell differentiation

1.2.1.1 Lineage-specifying transcription factors

The concept of distinct lineages has been reinforced by the discovery of lineage-specifying transcription factors for each subset, the expression of which is both necessary and sufficient for functional activity. These are known to bind and regulate key functional genes specific to their respective lineage. GATA-3, a member of the GATA family of zinc finger transcription factors, was found to be preferentially expressed in Th2 compared to Th1 cells [222] and GATA binding sites are present upstream of the IL4/IL5/IL13 gene cluster [223]. Knockdown showed GATA-3 was required for the expression of these Th2 cytokines [222], while ectopic expression of GATA-3, alongside stimulation, induces their expression [222]. A few

years after this discovery, a member of the T box family of transcription factors, T-bet (*T box expressed in T* cells, *Tbx21*) was found to be expressed in Th1 but not Th2 cells [224]. Other members of the Tbox family of transcription factors have roles in development. Members of this transcription factor family recognise a common sequence through a conserved DNA binding domain, the Tbox [225], and are recognised as both transcriptional activators and repressors [225]. Ectopic expression of T-bet activates *IFNG* [224] which contains a Tbox consensus binding site [224]. In addition, the loss of T-bet results in susceptibility to Type I infections [226]. Both GATA-3 and T-bet can re-polarise cells towards their respective lineages when expression is enforced in the alternative lineage [222, 224]. T-bet interacts with the H3K4 methyltransferase subunit RbBp5 and the H3K27 demethylase JMJD3 [227] and T-bet binding correlates with permissive histone modifications, including H3K9 acetylation [228] and H3K4 methylation [229], at Th1 upregulated genes. However as T-bet is also found at locations lacking these modifications and at loci that are not upregulated, it is unlikely these histone marks are a direct result of binding of the transcription factor [229]. While T-bet is known to be required for binding of the chromatin insulator CTCF (CCCTC-binding factor) [230], the biochemical mechanism of T-bet and T-box function in general is not fully understood.

In 2003, the publication of three studies demonstrated the requirement for the transcription factor FOXP3 for Treg function. This was elucidated through mapping the genetic cause of the murine autoimmune scurfy phenotype [37-39] and its human equivalent, IPEX syndrome [38] to the

FOXP3 locus. Supporting this, ectopic expression of FOXP3 conferred a Treg like suppressive activity to T responder cells [38] and limited autoimmune disease [39]. The equivalent transcription factor governing the Th17 lineage has since been identified as RORγt [231]. These transcription factors have historically been considered to be 'master regulators', a term that indicates omnipotence over cell phenotype. However, their co-expression has been frequently observed and fluidity between these lineages in now appreciated, leading to the more modest term 'lineage-specifying factors'. This is expanded upon in Section 1.2.2.

1.2.1.2 Lineage-plasticity and co-expression of lineage-specifying transcription factors

Although T cell lineages have distinct phenotypes, there is some plasticity between subsets. For example, Th17 cells can produce the Th1 cytokine IFNγ [232] while Treg may acquire a Th17 phenotype, which is pro-inflammatory rather than suppressive. This instability in Treg contributes to, rather than contains, inflammation and has been implicated in the pathogenesis of autoimmunity [233, 234]. Genetic lineage tracing has been used to track the fate of mouse Treg *in vivo*. This showed that ex-Treg, which had downregulated Foxp3*,* were increased in inflamed tissues and their adoptive transfer promoted autoimmune diabetes [235]. Such instabilities in T cell lineages also have functional rather than pathogenic consequences. For example, fully differentiated Th2 cells can repolarise *in vivo* to a Th1 functionality when confronted with a Type I infection [236]. Expression of the lineage-specifying transcription factors reflects this diffusion of T cell identity. For example, Foxp3 is seen in activated but non-suppressive T effectors and may reflect chromatin remodelling of the

Foxp3 locus downstream of TCR signalling and may be expressed without the associated Treg phenotype [235]. In addition, the expression of these transcription factors is not mutually exclusive as once thought and co-expression can be seen in the same cells [237]. This co-expression may have functional importance. The expression of T-bet [238-241] or Gata3 [242] alongside Foxp3 in Treg may allow Treg to co-opt a partial Th1 or Th2 phenotype, respectively, for more targeted regulation. Supporting this theory, T-bet expression in FOXP3-expressing Treg occurs in response to Th1 cytokines IFNγ [239] and IL-27 [241] and leads to expression of Th1 associated chemokine receptor CXCR3 [239], with accumulation of T-bet positive Treg being seen during type 1 inflammation [239]. In addition, Gata3 complexes with Foxp3 and its deficiency leads to Th2 mediated inflammatory disease [242].

Beyond Treg, the repolarisation of Th2 cells to functional Th1 cells, mentioned above, is reflected in transcription factor expression [236]. Work from our lab and others [243, 244] has shown that Gata3 is also expressed in Th1 cells [236, 243-246] but that expression of T-bet has a dominant effect [245, 246]. The presence of Gata3 in Th1 cells may be functional in controlling expression of pan-T cell active genes, such as TCR genes [243]. In addition, expression of T-bet and GATA-3 has been observed in RORγt expressing Th17 memory cells, with a suggestion that this may allow potential for repolarisation [247], which has been demonstrated through priming of Th1 cells with the Th17 type pathogen, *Candida albicans* [248]. Co-expression has also been observed in several immunopathologies such as experimental allergic encephalomyelitis [249],

juvenile idiopathic arthritis [250] and allergic asthma [251] and may contribute to pathology.

Our lab has explored the interplay between T-bet and GATA3 with ChIP-Seq demonstrating that GATA3 binding transitions from Th2-specific sites in Th2 cells to T-bet Th1-specific sites in Th1 cells [246], suggesting that T-bet acts dominantly in the relocation of other transcription factors. This is consistent with suggestions that T-bet's primary function is to repress the alternative Th2 transcriptional program [252, 253]. A similar relationship has been seen between TBX5 and GATA4 during cardiac development [254], suggesting a common mechanism for T-box family member transcription factors. Furthermore, T-bet binds the Bcl6 transcription factor DNA binding domain thereby occluding its function [255, 256]. These observations show that lineage-specifying transcription factors are not omnipotent and that defining T cell subsets on the basis of their expression alone is limited. Therefore it is important to understand the additional levels of regulation involved not only at lineage specific genes but genome wide.

1.2.1.3 STATs and other pioneer factors

As described above, TCR signalling initiates activation of activator protein (AP) -1, NFAT and NF-κB. These transcription factors are not lineage-specific [257] but their binding precedes and theoretically enables recruitment of lineage-specifying transcription factors [258]. For example, FOXP3 appears to primarily bind sites targeted by these transcription factors as they are already accessible in Treg precursor cells [258]. Such precursor transcription factors that allow the activity of lineage-specifying factors have been termed 'pioneers' [258]. NFAT [259, 260] and NF-κB

[259] have been shown to cooperate with FOXP3. Similarly, cofactors have been postulated to pre-bind DNA, facilitating open chromatin conformation prior to FOXP3 binding [258]. In addition, transcription factors of the same structural family may act as placeholders, for example FOXO1 through its similarity in binding specificity to FOXP3 [258].

The STAT transcription factors are also considered to be pioneers and are activated in a lineage-specific manner [261]. STAT activation occurs downstream of cytokine signalling. When IL-12 binds its receptor, STAT4, sequestered at the membrane is phosphorylated and enters the nucleus, where it binds and activates expression of the Th1 signature cytokine IFNγ [262-265]. This in turn has an autocrine effect by activating STAT1 [266]. IL-4 meanwhile, activates STAT6 [267], culminating in expression of the Th2 lineage-specifying transcription factor GATA-3 [268]. STAT3 is activated by IL-6 [269] leading to Th17 differentiation [270] and IL-2 activates STAT5 for Treg differentiation [271]. Binding site motifs for these STATs are present in the regulatory regions of the genes encoding the lineage-specifying transcription factors. For example, STAT4 and STAT1 bind the gamma-interferon-activation-site-with-three-base-pair-spacer (GAS-3) motif upstream of *IFNG* [261] and STAT6 binds GAS-4 upstream *of IL4* [261]. Genome-wide profiling shows that STATs are extensively associated with regulatory regions, even to a greater extent than lineage-specifying factors [261] and deficiency of STATs leads to disruption of these regulatory regions, which cannot be rescued by expression of lineage-specifying transcription factors [261].

1.2.1.4 Cofactors

The biochemical mechanisms of how the T cell transcription factors regulate gene expression are not well understood but given their binding is associated with both up and downregulation of gene expression [272, 273], their specific modes of activity are likely influenced by binding partners. In addition to augmenting the activity of transcription factors, cofactors have potential to act as placeholders at regulatory regions thereby directing transcription factor binding.

T-bet cofactors include ETS1 [274] and RUNX3 [275]. Interaction of T-bet with RUNX3 competes with Gata3 and is necessary for repression of Il4 through this mechanism [276, 277]. The presence of ETS1 binding upstream of *Ifng* is necessary for T-bet mediated induction of *Ifng* and type 1 inflammation *in vivo* [274]. A mass spectrometric analysis identified over 300 proteins associated with FOXP3 that may act to regulate its activity [242]. This included transcription factors that were already known to have roles in Treg such as FOXP1, RUNX1 and IKZF1 (Eos), as well as components of several chromatin-modifying complexes [242]. While these cofactors are not Treg specific, synergy with FOXP3 may target their activity to Treg relevant genes. The Th2 lineage-specifying GATA3 was also found in this FOXP3 'interactome' complex [242] and is also known to bind the FOXP3 promoter [278], and vice versa [242]. In *Foxp3*^{Cre}*Gata3*$^{fl/fl}$ mice, although the number of Treg was similar to littermate controls, the Treg-specific loss of Gata3 was shown to increase Th2 mediated intestinal inflammation and dermatitis [242], demonstrating that this interaction is immunologically relevant. This co-expression of FOXP3 and an

inflammatory lineage-specifying transcription factor may allow Treg to adapt their phenotype to the immunological setting, for example through the expression of Th1 or Th2 homing receptors, allowing their targeting to relevant anatomical sites [239].

1.2.2 Enhancers and regulatory regions

Enhancer regions are DNA elements that increase expression of a gene above the level seen with the minimal promoter region [279], can exert their influence over various distances [280], regardless of orientation [279] and can recruit RNA polymerase II (RNAPII) in the absence of a promoter [280]. These characteristics, first described in reference to the enhancing effect of enforced expression of the SV40 promoter on the beta-globin gene, have come to form the basic definition of the enhancer element in metazoans [281]. Following this, the first intrinsic eukaryotic enhancer to be described was at the immunoglobulin heavy chain (IgH) locus, where chromosomal rearrangements outside of the promoter-proximal region are required for high level expression of the recombined immunoglobulin gene [282-284]. These properties ultimately lead to a communication between enhancers and promoters that regulates the expression of a gene. This communication can be enhanced with increasing proximity between the enhancer and promoter [279] but can also occur over large genomic distances [285].

1.2.2.1 Features of enhancers

In addition to the properties described above, additional features of enhancers include an association with DNaseI hypersensitivity sites (DHS), the binding of transcription factors and coactivators and the

presence of signature histone modifications. DHS are identifiable through nuclease treatment followed by DNA sequencing and represent genomic regions with enhanced accessibility, as indicated by increased cleavage [286]. This accessibility is directly related to transcriptional activity through an increased capability to bind regulatory complexes. Supporting this, DHS loci are associated with functionally relevant and cell type specific genes [287]. The profiling of genome wide DHSs has been used to identify *de novo* regulatory regions, including putative enhancers [258, 288]. However, DHS are also found associated with genes of closely related alternative cell fates and so, for example, naïve T cells and Treg share 99% of DHS sites [258]. In addition, not all genes associated with DHSs are transcribed [289]. These observations have lead to a distinction between active enhancers, which are currently operating within a cell to activate gene expression, and inactive but poised enhancers, the accessibility of which allows a response to the cellular environment should an alternative cell fate or transcriptional program be required [290]. Both active and poised enhancers are marked by mono-methylation of histone H3 at lysine 4 (H3K4me1, Section 1.2.3, Table 1.1, [289]), a permissive histone modification that is associated with accessible areas of chromatin and is lost as enhancers are decommissioned during differentiation [291]. However, only active enhancers are marked by acetylation of H3K27 [290], a mark that is gained at lineage-specific loci as differentiation progresses [290]. It is not fully understood if these histone modifications are a cause or consequence of enhancer formation and activity, however their deposition has electrostatic effects on chromatin conformation and provides binding sites for regulatory proteins (Section 1.2.3, Table 1.1). In addition, the

binding of transcription factors, coactivators and chromatin-modifying complexes with known activities is informative of enhancer formation. For example, the coactivator CREB binding protein (CBP) and its family member P300 have histone acetyltransferase activity [292] and their genomic occupancy corresponds with DHS sites and has been used to identify active enhancers [293-295]. The transcriptional coactivator BRD4 (Bromo-domain containing 4) [296] recognises acetylated histones, such as H3K27ac [297], is required at active genes for processive elongation of transcription (Section 1.2.2) and has been associated with active enhancers [296, 298]. The co-assembly of multiple complexes and their cooperation at enhancers is recognised and termed 'The Enhanceosome' [299] and binding site motifs for transcription factors are also found at enhancers.

Supporting the fundamental importance of enhancers, several diseases are associated with disruptions of enhancer relevant complexes. This includes general developmental abnormalities seen in Rubinstein-Taybi syndrome caused by CBP or P300 mutations [300, 301] and in Cornelia de Lange syndrome, which is associated with mutations in components of the Cohesin complex [302, 303]. Additionally, mutation of enhancer sequences but not coding sequences are involved in several diseases including Burkitt's lymphoma, in which chromosomal translocations have been observed in close proximity to the oncogenic MYC gene [304], and a subset of β-thalassemias that show disruption of an upstream DNaseI hypersensitive region [305].

1.2.2.2 Mechanisms of enhancer function

The communication between enhancer and promoter for the influence of gene expression is exemplified in Drosophila by insertion of *gypsy* retrotransposon element, which blocks enhancer-promoter interaction through recruitment of the chromatin insulator protein, Hairy wing [306]. In mammalian cells this formation of chromatin boundaries between enhancers and promoters is mediated by CTCF (CCCTC-binding factor) [307]. Potential mechanisms for enhancer-promoter communication include: tracking of transcriptional machinery from an enhancer to a promoter, looping of DNA to bring enhancer and promoter into close proximity independently of the intervening DNA, and cooperation of both mechanisms [281]. Theories of tracking suggest complexes are recruited to an enhancer, they may migrate along the DNA until reaching the promoter [308]. Looping of DNA to allow interaction between an enhancer and promoter without interaction with the intervening DNA sequence is supported by the chromosome conformation capture technique, which identifies physical interactions between genomic regions. This looping has been demonstrated at the beta-globin locus with mediation through CTCF [309]. The conserved non-coding sequences that CTCF binds have also been shown to interact with Cohesin complexes [310], which have a canonical role in sister-chromatid cohesion [311] but have also been shown to influence gene expression [312]. Mediator, a macromolecular coactivator complex that binds transcription factors at regulatory regions and promoters, also influences gene expression through DNA looping in cooperation with Cohesin [313, 314]. Complications in predicting the communication between enhancers and promoters arise from the ability of

enhancers to communicate with multiple promoters [315] and multiple promoters to contact a single enhancer [316]. In addition, while a shorter distance between an enhancer and promoter can increase the effect on gene expression [279], long range communications are also efficient in other circumstances [285] and so relative position alone does not predict interaction [317].

1.2.2.3 Super-enhancers

Locus control regions (LCRs) are defined by their ability to enhance the expression of linked genes to physiological levels in a tissue-specific and copy number dependent manner at ectopic chromatin sites [318]. Genomic methodology has allowed similar regulatory elements to be identified across the genome, as highly occupied target (HOT) regions, transcription initiation platforms (TIPs, [319]) or 'super-enhancers' [320, 321]. HOT loci have been defined in Drosophila as regions that exhibit enriched binding for multiple transcription factors [322, 323] with this increased transcription factor binding complexity correlating with nucleosome displacement [324], a feature of active chromatin. Compared to classical enhancers, this accumulation of multiple transcription factors with modest individual influence may provide critical mass allowing a synergistic effect on gene expression [322]. These binding patterns can be used as a predictor of the *in vivo* functional effect of transcription factors for which there is limited understanding [325]. In mouse similar regions with high transcription factor occupancy have been referred to as transcription initiation platforms (TIPs) [319]. This transcription factor density has been shown In ESCs through binding of the pluripotency master regulator transcription factors Oct4, Sox2 and Nanog [321] and interpreted as clustering of enhancers to form

stretched regions of ‘super-enhancers’ [320, 321]. These super-enhancers exhibit preferential binding of the transcriptional coactivator BRD4 and are associated with oncogene transcription [320]. Following this, pharmacological inhibition of BRD4 leads to disruption of binding preferentially at super-enhancers with a decrease in transcription from the associated oncogenes [320]. Supporting a role in disease, super-enhancers are enriched for disease associated DNA sequence variants in disease relevant cell types [326]. Compared to normal enhancers these clusters exhibit increased levels of the active enhancer marks of H3K27Ac and DNaseI hypersensitivity and correlate with greater transcription from associated genes with binding of transcription factors being necessary for this. This transcription from associated genes is also cell type specific in differentiated cells of the haematopoietic system [321].

1.2.2.4 T cell enhancers and regulatory regions

Demonstrating the importance of transcription factor binding at enhancers, T cell enhancers are enriched for lineage relevant STAT binding sites with the STAT1 and STAT4 motifs corresponding to Th1-specific enhancers and the STAT6 motif corresponding to Th2-specific enhancers [261], while general T cell transcription factor motifs such as AP-1 and NFAT, which are activated following TCR signalling, are found at both [261, 319]. Such transcription factors are activated downstream of signalling pathways and so this enrichment of binding demonstrates the link between environmental stimulus and enhancer activity through the priming of enhancers through chromatin remodelling [258, 261]. Furthermore the binding of P300 is dependent upon these lineage-associated transcription factors [261, 327]. A recent profile of the histone and transcriptional landscape in human Tresp and functional expanded Treg identified differential enrichment of transcription factor binding motifs at enhancers. This included greater enrichment of KLF (Krüppel-like factor), FOX (forkhead) and STAT5 motifs at Treg enhancers and greater enrichment of RUNX motifs at Tresp enhancers, with ChIP-Seq binding data reflecting this motif enrichment [328]. Co-factors of lineage-specifying transcription factors and structural family members have been shown to have a role at poised enhancers [258]. For example, enhancers that are bound by the Treg lineage-specifying transcription factor FOXP3 in Treg are accessible and bound by FOXP3 cofactors, including ETS and RUNX, in Treg precursors, which may facilitate FOXP3 recruitment [258]. FOXO1 (forkhead box protein O1), which does not show Treg specific expression but is structurally homologous to FOXP3 and is essential for Treg development [329, 330],

may act as a pioneer, binding the forkhead motif at enhancers in Treg precursors and being displaced by FOXP3 upon Treg differentiation, demonstrating another mechanism for maintaining accessibility of enhancers prior to their activation [258].

1.2.2.5 IFNG – a paradigm for lineage-specific gene regulation

The *IFNG* regulatory region is relatively well characterised and as such represents a paradigm for the study of gene regulation and enhancer function. The characterisation of this regulatory region began with the identification of several DHS sites in Th1 cells [331] with transgenic luciferase reporter assays [332-334] and deletion experiments [335, 336] allowing identification of the minimal region required for optimal IFNG expression. Nomenclature of the individual enhancer regions has been based on the location relative to IFNG of several conserved non-coding sequences (CNSs), to which DHS sites map [335, 337-340]. The occupancy of transcription factors including GATA3, several STATs [264, 341-343] and of course T-bet [335] have now been profiled as have the presence of permissive and repressive histone modifications [344-346]. In addition, the influence of 3D chromatin architecture has also been explored at IFNG through binding of the chromatin insulator CTCF [230] and cohesin [347].

1.2.3 Post-translation histone modifications

Histone proteins are the basis for physical organisation of DNA within the nucleus and thus are closely related to the accessibility of DNA and gene expression. Four families of histone protein, H2A, H2B, H3 and H4, exist

as homodimers in an octomeric complex which together with the linker histone H1 forms the nucleosome around which 146 base pairs of DNA is wrapped. The basic nature of histone proteins facilitates their interaction with the negatively charged DNA phosphate backbone and modification of this electrochemical interaction directly impacts upon accessibility of DNA to other nuclear regulatory components. These modifications are covalent and chemically varied, and include the deposition of small chemical groups through acetylation ($COCH_3$), methylation (CH_3) and phosphorylation (P). Larger moieties can also be added through ubiquitinylation (addition of a ubiquitin polypeptide), SUMOylation (small-ubiquitin-like-modifier proteins) and ADP-ribosylation (ADP-linked ribose sugar). These modifications frequently occur on the tails of histone, which protrude out of the nucleosome. Nomenclature for histone modifications (eg. H3K4me3) is based upon the histone protein (eg. H3), the residue of this protein (eg. K4) and the modification (eg. trimethylation – me3).

As well as influencing DNA accessibility through electrochemical alteration, these chemical modifications create binding sites for additional chromatin modifiers and regulatory molecules with more specific activities. As histone modifications occur in various combinations and within the context of other chromatin modifications, their direct biological causes and consequences are not fully understood. This understanding is further complicated by the general topology of chromatin and potential for positional or sequence context effects. However some have clear associations with open and active chromatin while others are found at repressed chromatin. Those relevant to this work are described in Table 1.1. Molecules involved in the

deposition (writers) and recognition (readers) of these modifications are also detailed.

1.2.3.1.1 Acetylation

With the addition of a negatively charged acetyl group, the positive charge of histones is neutralised leading to a reduced affinity of the histone to the negatively charged DNA, resulting in greater conformational flexibility and accessibility. Acetylation of histones is therefore broadly associated with transcriptionally active regions of chromatin. Histone acetyl transferases include the P300 and CREB binding protein (CBP) family of co-activators [292]. In addition to their electrostatic influence, acetylated lysines, such as H3K27Ac, are recognised by bromodomains [348] including that in BRD4 (Bromo-domain containing 4) [297], a complex that positively regulates transcription [349]. Due to their functional relevance, the occupancy of CBP/P300 [295] and BRD4 [281, 350] has been used to identify active, as opposed to poised, enhancers [290].

1.2.3.1.2 Methylation

Methylation of histones is associated with activation and repression, depending upon context, with deposition being catalysed by histone methyltransferases, of which several can target the same residue [351]. For example H3K4 is modified by several MLL and SET complexes [351]. Tri-methylation of histone H3 at lysine 4 (H3K4me3) is found at the 5' end of actively transcribed genes in association with the transcription initiating form of RNAPII [351-353], which is phosphorylated at Ser5 of the C-terminal domain. In contrast H3K36me3, catalysed by the histone methyl transferase Set2 [354], occurs along the gene body, accumulating towards

the 3' end, and is associated with the transcription elongating form of RNAPII (phosphorylated at Ser2 of the C-terminal domain) [351, 355]. H3K4me1 can be found flanking H3K4me3 at active and poised genes [356] and, in addition, is associated with enhancer regions, although there is debate over the whether in this context it occurs in the presence [344] or absence [293] of H3K4me3. H3K27me3, meanwhile, is a feature of the 5' end of silent genes residing in open chromatin and is deposited by the histone methyltransferase (HMT) EZH2, a subunit of polycomb repressive complex 2 (PRC2) [357]. H3K27me3 is present at *IFNG* and *TBX21* and other Th1 genes in Th2 cells and is also present at *IL4/5/13* and *GATA3* in Th1 cells [340, 358]. In addition, PRC2 is necessary for Th1-cell specification as demonstrated in *Ezh2* deficient mice in which the repression of alternative lineage genes including *Tbx21* and *Gata3* is dysregulated leading to a Th2-mediated asthma like pathology [359].

H3K9me3 has been found in large areas of repressed chromatin where its deposition is catalysed by the Suv39 class of HMTs [360] and it is associated with transcriptional repression of potentially deleterious transposons and repeats at pericentromeres [361, 362]. In addition, it has also been reported along the body of elongated genes [363].

Histones are methylated during transcriptional processes and create binding sites for complexes that function in later stages of transcription. Methylation is recognised by highly conserved chromodomains, which are present in several chromatin-modifying complexes with varying specificities. For example, chromodomain containing HP1 recognises

H3K9 methylation [364] and CBX subunits of PRC1 recognise H3K27 methylation [365]. Both of these recognition events are related to subsequent transcriptional repression. H3K36me3 is recognised by chromodomain containing Eaf3, part of an HDAC-containing complex, which functions in transcriptional elongation through the repression of cryptic start sites in the gene body post-expression [355]. Aside from this, H3K4me3 is recognised by the NURF complex [351] which catalyses ATP-dependent chromosome remodelling [366] allowing transcriptional activation [367]. H3K4me1 is recognised by the WDR5 subunit of the MLL1 complex [351, 368] although the functionally significant events downstream of this recognition are not fully understood [351].

Combinatorial modifications include the active mark H3K4me3 and the repressive mark H3K27me3, which are paradoxically found together in bivalency [369]. This is seen at tissue specific genes of alternative cell fates in embryonic stem cells, which become de-repressed upon differentiation [369]. This has been suggested to allow the poising of lineage-specific genes for their rapid activation in future cell states. The same phenomenon also occurs during adult T cell differentiation where in naïve cell precursors both histone marks are found at *Tbx21* and *Gata3* but H3K27me3 is lost at *Tbx21* in Th1 cells and at *Gata3* in Th2 cells [358]. This demonstrates that genes corresponding to alterative fates are not fully switched off but instead remain poised for activation in the future, should they be required, which may be important for lineage plasticity as described in Section 1.2.1.2

Table 1.1: Example histone modifications and their associated transcriptional status

The post-translational modification of histones is associated with both activation and repression of transcription. These modifications are deposited by histone modifying complexes (writers) and recognised by adapter molecules and chromatin-modifying complexes (readers). Brackets detail complexes in which these enzymatic subunits are found. The details of histone modifications and associated readers and writers given here are limited to those with relevance to this thesis.

Histone	***Residue***	***Modification***	***Writers***	***Readers***	***Transcriptional association***
H3	K4	Mono-methylation (me1)	MLL1-5, SET1A, B [351]	WDR5 (MLL1) [368]	Found at enhancers [293]. Also seen at active promoters in combination with H3K4me3 [356].
		Tri-methylation (me3)	MLL1-5, SET1A,B [351]	BPTF (NURF) [351, 367]	Associates with the initiating form of RNAPII and found at start sites of active and poised genes in yeast [352] and higher eukaryotes [353].
	K27	Tri-methylation (me3)	EZH2 (PRC2) [357]	CBX7 (PRC1) [370]	At inactive genes and (in association with K4me3) at developmentally regulated genes poised for activation [357, 369]
		Acetylation (ac)	P300-CBP family [292]	Bromodomains [348] including BRD4 [297]	Presence in combination with H3K4me1 marks active, as opposed to inactive or poised, enhancers [290].
	K36	Tri-methylation (me3)	SET2 [354]	Eaf3 [355]	Transcriptional elongation. Presence along the gene body, in conjunction with H3K4me3 at the promoter region, indicates active and processive transcription [355].
	K9	Tri-methylation (me3)	Suv39h HMTs in heterochromatin [360], EHNT HMTs in actively transcribed genes [371]	HP1 [364]	Found in repressed heterochromatin but also associated with transcriptional elongation in actively transcribed genes [363] and at active promoters [372].

1.2.3.2 Transcriptional elongation by P-TEFb

Effective regulation of transcriptional elongation in addition to initiation is necessary for gene expression. As described in above, at lineage specific genes of alternative cell fates the presence of histone modifications associated with transcriptional initiation in bivalency with those associated with repression demonstrates the importance of transcriptional elongation for lineage choice.

RNAPII recruitment to the core promoter results from the assembly of a cascade of basal transcription factors following the initial binding of the TATA binding protein (TBP), which commonly recognises the TATA box sequence motif [373]. This RNAPII recruitment is stochastic and dependent upon accessibility of the core promoter, which is influenced by chromatin conformation. Therefore RNAPII recruitment may occur at promoters that are accessible but not relevant to the cell's current state. Regulation at the point of elongation therefore enforces cell type specific gene expression. Initial RNAPII binding leads to the production of short transcripts of 20-60nt due to pausing adjacent to the promoter [374, 375]. For progression of RNAPII out of the promoter-proximal region, and for productive elongation of transcription through the gene body to occur, the carboxyl terminal domain (CTD) of RNAPII must be phosphorylated at the serine 2 position. Conversely, paused RNAPII is phosphorylated at the serine 5 position. This phosphorylation is catalysed by the positive transcription elongation factor (P-TEFb). P-TEFb is a heterodimeric complex including CyclinT1 (or T2) and the cyclin dependent kinase CDK9 [376]. P-TEFb activity is negatively regulated through sequestration by the

scaffolding RNA 7SK [377, 378] and is recruited in its active form to promoters by BRD4 [349], which recognises acetylated histones, or as part of the super-elongation complex [379]. Both the inhibition of CDK9 kinase activity by Flavopiridol [380] and the blocking of the acetyl-recognising bromo and extra-terminal (BET) domain of BRD4 by JQ1 [381] lead to genome-wide inhibition of transcription. In addition, BRD4 occupancy has been identified at enhancers [296, 298, 320, 326, 382-384] but the importance of P-TEFb here is not well understood. Some reports have demonstrated that P-TEFb activity is required for downstream gene expression [385] while others challenge its role at enhancers [386, 387]. So far, P-TEFb occupancy at super-enhancers has not been observed.

1.3 Non-coding RNA

1.3.1 Refining the central dogma of molecular biology

The understanding of the basis of molecular biology has traditionally placed RNA in the role of regulating translation of DNA-encoded instructions into proteins, the effectors of cell phenotype. In addition to messenger RNAs (mRNA), this includes ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), which are denoted by a U- prefix descriptive of their high uridine content, have a central role co-transcriptional RNA splicing (reviewed in [388]), and small nucleolar (snoRNA), which are located in the nucleolus where they have a canonical role in ribosome synthesis and the post-transcriptional modification of RNA [389]. Supplementary to these roles, microRNAs (miRNA) regulate translation through the targeting of complementary mRNA transcripts for destruction or by inhibiting their translation [390]. Hinting at translation-independent roles for small RNAs, piwi-interacting RNA (piRNA), so named for their processing by piwi proteins, which are gamete specific members miRNA biogenesis machinery, were identified as having a role in transcriptional silencing of repetitive regions such as transposon sequences during spermatogenesis thereby promoting genomic fidelity [391]. In 2001, the sequencing of the human genome unexpectedly uncovered that the number of protein-coding genes does not increase with species complexity, with our genome committing less than 2% of its sequence to this task [392]. The advent of whole genome sequencing techniques has demonstrated that the majority of the genome, while not coding for protein, is transcribed into RNA [393]. Suggestive of functional importance, this pervasive transcription has been shown to be at least in

part tissue specific and developmentally regulated [394] and subject to RNA processing mechanisms including polyadenylation and splicing [395]. The role of non-coding DNA and the non-coding (ncRNA) produced from it has been debated with the assertion that transcriptional activity of 'junk DNA' does not equate to function [396, 397]. Transcription of ncRNA could be a stochastic by-product of unregulated accessibility of DNA to transcriptional machinery. In addition, functional consequences can lie in the act rather than the product of transcription and, where the production of RNA is functional, the sequence identity of the RNA itself may be irrelevant. However some ncRNAs have been demonstrated to be functional. This is exemplified through the requirement of the ncRNA *XIST* for X-chromosome inactivation [398]. Several other ncRNAs have annotated functions, some related to lineage specification, and are described below.

1.3.2 Genomic classification of non-coding RNA

The classification of these newly identified transcripts on the basis of location of genomic origin and size of transcript has been fundamental to understanding the consequences of the transcription of ncRNAs [399] and has been facilitated by current sequencing technologies. On the basis of purification techniques, ncRNAs have been crudely divided into short and long, 200 nucleotides being the cut-off length [400, 401]. Nether of these fractions are homogenous in identity, containing transcripts that are diverse in properties and function. These properties influence their regulation, activity and function, which lead to further classification of transcripts. Mechanistic details have been delineated for several members of these classes and are discussed in Section 1.3.3**.** These mechanisms

are not common to all transcripts with the same genomic classification and may be shared with transcripts with dissimilar genomic origins. Also important to note is that some presumed ncRNAs may be found to code for short peptides and so the ultimate proof of non-coding function is in demonstrating mechanistic importance of the transcript itself [402].

Figure 1.4 illustrates how the genomic location and structure of ncRNAs influences their identity and how several classes have been defined on this basis and have potential roles in transcriptional regulation of lineage specific gene expression. These classes, as described in Sections 1.3.2.1 to 1.3.2.4, include long intergenic ncRNAs (lincRNAs), natural antisense transcripts (NATs), promoter associated short RNAs and enhancer RNAs (eRNAs).

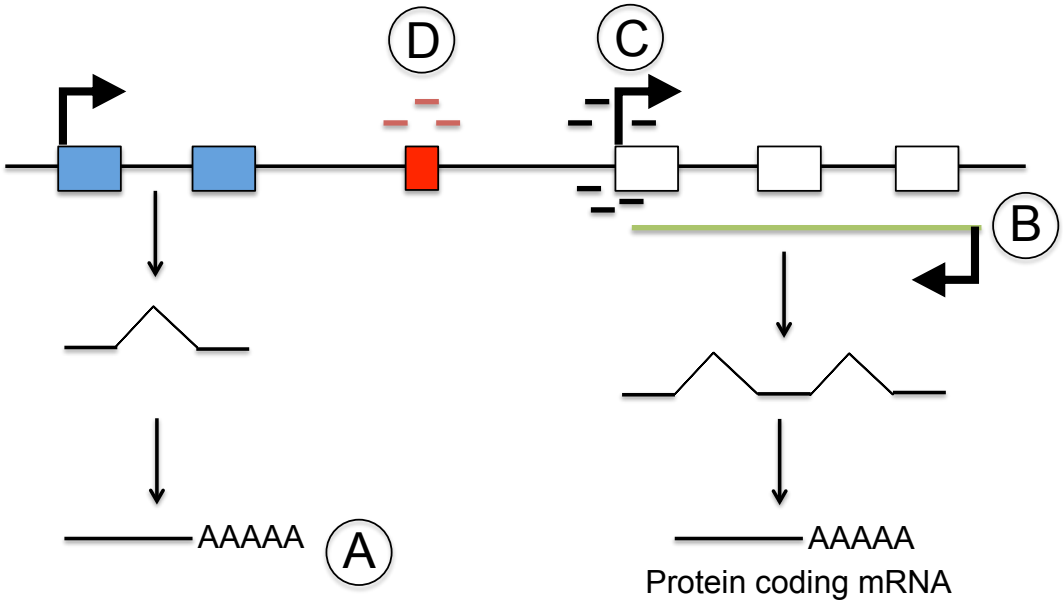


Figure 1.4: Classes of ncRNA in relation to genomic origin

The genomic location and structure of ncRNAs influences their identity and ultimately function. A: lincRNAs are transcribed, spliced and polyadenylated similar to mRNAs. B: Natural antisense transcripts are transcribed antisense to protein coding genes. C: Short non-coding transcripts are transcribed from transcription start sites. D: NcRNAs transcribed from enhancer regions may mediate the long range effects of enhancers.

1.3.2.1 Long intergenic non-coding RNAs

Many long ncRNAs are encoded in a genic structure similar to protein-coding genes in that they have distinct introns and exons, can be alternatively spliced under canonical splice site signals [403] and are processed in a similar manner through capping and polyadenylation [404]. These have been termed long intergenic ncRNAs (lincRNAs), reflecting their position in intergenic regions, as shown in Figure 1.4.A. LincRNA loci are now recognised as genes themselves, just lacking open reading frames longer than 100 codons [402], and they are catalogued by annotation consortia such as RefSeq and Gencode [403]. LincRNA genes are marked by the same histone modifications as protein-coding genes, which are reflective of transcription status, including H3K4me3 and H3K36me3 at the TSS and gene body of transcriptionally active genes, respectively [405]. Unlike protein-coding genes, lincRNA genes show preference for comprising two exons [403], although some are single exon (eg. *Malat1* and *Neat1)* [404] and their exons and introns are longer [403]. LincRNA genes also show enrichment of sequences derived from common repeats, including LINES, SINES and LTRs [403, 406, 407]. These repeat sequences may be key to influencing secondary structure, for example through self annealing to form stem-loop structures, which may form binding sites for regulatory protein complexes. However, not all lincRNAs contain repeats [403]. A key distinction between protein-coding and lincRNA genes is in sequence conservation; non-coding sequences show less evolutionary conservation than protein coding sequences as they are under different selective pressures and evolutionary constraints as they are not restricted by codon choice required to maintain protein structure

and function. However, lincRNAs exhibit more selection pressure than neutrally evolving sequences [403-406] and exons and promoters are more conversed than introns [403, 405]. Interestingly, although expressed at a lower level [408], lincRNAs show more tissue specific expression than protein coding genes [403]. The importance of some lincRNAs has been challenged through observations that loss of transcript experiments often have no functional consequences, for example HOTAIR [409]. This may reflect a specificity of function for these transcripts, which perhaps is only exhibited under certain conditions and organismal or cellular pressures or at specific temporal points of differentiation. However, in support of lincRNA function, a recently developed cohort of lincRNA knockout mice has demonstrated the functional importance of several lincRNAs [410]. LincRNAs for which knockout produced extreme effects included *Peril*, for which knockout resulted in perinatal lethality (although with only 50% penetrance for the homozygote), possibly due to associated changes in gene expression within the brain, Fendrr, the loss of which also resulted in death soon after birth, which was associated with respiratory failure and oesophageal and gut abnormalities, and *Mdgt*, the loss of which leads to stunted growth [410].

1.3.2.2 Natural antisense transcripts

Natural antisense transcripts (Figure 1.4.B) by definition are encoded antisense to other genes. They may or may not share the genomic features of lincRNA such as canonical start and stop sites, splice sites and polyadenylation signals. Their genomic orientation is often relevant to a biological role through influencing transcription of the sense-strand gene. This influence can be positive through recruitment of transcriptional

machinery or negative through transcriptional interference. These mechanisms are elaborated on in Section 1.3.3.1.

1.3.2.3 Promoter-associated RNAs

Transcripts of heterogeneous lengths, are produced from transcription start sites (TSS) of genes in addition to the production of full length, polyadenylated RNA (Figure 1.4.C) [411-415]. Transcripts of 50-200nt have been termed promoter-associated short RNAs (PASRs) [411] or TSS-associated RNAs (TSSa RNA) [412], while smaller transcripts with a similar distribution around the TSS have been termed tiny RNAs [415]. These transcripts are produced bidirectionally, transcribed in both the same and divergent orientations to the host gene [412, 413, 415], and are associated with paused RNAPII [416]. Our lab has shown that RNA of 50-200nt are transcribed from the TSS of polycomb target genes in primary human T cells with their presence being independent of polycomb activity [416]. These RNAs interact with PRC2 and are lost from polycomb target genes that are activated upon cell differentiation, indicating their presence may support the interaction of polycomb complexes with chromatin at these genes [416]. Transcripts described elsewhere have been termed promoter upstream transcripts (PROMPTs). PROMPTs have been characterised as short, polyadenylated and capped transcripts that are ubiquitously transcribed bidirectionally from 0.5 to 2.5kb upstream of active TSSs [417]. These transcripts have been suggested to arise from stalling of transcriptional machinery [418]. The transcription of promoter associated short RNAs may promote the maintenance of chromatin in an open conformation at a TSS and may provide a reservoir of transcriptional machinery in close vicinity for rapid induction of gene expression [419].

1.3.2.4 Enhancer RNAs

RNAPII binding has historically been observed at active enhancers [280] and has more recently been shown to be functional through the transcription of RNA at these sites [386, 420]. The transcripts produced from enhancers can be polyadenylated or not [420, 421], may be bidirectionally transcribed [420-423] and are predominantly unspliced and unstable [386]. These eRNAs are low abundance compared to mRNAs and lincRNAs, an observation that is explained by their low stability [386] and rapid degradation by the exosome, the macromolecular complex that catalyses RNA decay**.** It has been suggested that this degradation may lead to the production of potentially functional shorter transcripts.

In neuronal cells, Kim et al (2010) showed that transcriptionally active enhancers, identified by binding of the transcriptional coactivator CBP, are marked by H3K27ac and H3K4me1 [420]. Compared to RNA produced from promoters, these transcribed regions showed a lower ratio of H3K4me3 to H3K4me1. Also in contrast to promoters, which can exhibit bidirectional transcription but with preference for sense strand production, these enhancers RNAs showed no bias in directionality.

Theories for a transcription independent role for RNAPII processivity at enhancers include tracking of the transcriptional machinery towards a promoter (Section 1.2.2.2 [308]). It has also been suggested that the transcriptional activity that produces eRNAs may function to keep an enhancer region in an open chromatin confirmation by steric force or through maintaining separation of DNA strands thereby enforcing

accessibility [424]. In addition, the process of transcription is required for histone acetylation at extra-genic loci, as shown by the effect of inhibition of transcription with actinomycinD, which is not seen with inhibition of protein synthesis with cyclohexamide treatment [386], and may allow for enhancer functionality.

Suggesting a function for this transcription, the level of eRNA frequently correlates with mRNA transcription from nearby genes [420]. In macrophages**,** 70% of genome wide RNAPII binding is associated with enhancers and results in transcription in response to stimulation by endotoxin which is temporally coordinated with proximally located protein-coding genes [386], suggesting a role in environmental response. The production and dispersal of eRNA (Figure 1.4.D) may provide a mechanism for the long range effects of enhancers, which can be distantly located to the genes they regulate. These transcripts may aid maintenance of enhancer-promoter interaction through looping [425-427] or allow the targeted recruitment of chromatin-modifying complexes and transcriptional machinery, explaining how these complexes are targeting to specific genes. Functional experiments have demonstrated that the presence of eRNA transcripts themselves are required for optimal expression of their associated genes [420, 421, 427-429]. Mousavi et al (2013) examined eRNAs transcribed from muscle cell enhancers bound by the myogenic transcription factors MyoD and MyoG. Knockdown with siRNA of transcripts from a core enhancer marked by H3K4me1 upstream of MYOD1 lead to reduced MYOD1 expression at both the transcript and protein level. In addition, knockdown and overexpression of a transcript

from a distal regulatory region affected MYOG expression [428]. In macrophages, Lam et al (2013) showed that the repressive nuclear receptor Rev-Erbs inhibits gene expression by targeting enhancers that are suppressed in a macrophage specific manner [429]. The function of Rev-Erbs in this context was associated with inhibition of transcription from these enhancers [429]. Again, assays for knockdown of eRNAs demonstrated a corresponding effect on nearby mRNA production [429]. During LPS-induced monocyte activation, Ilott et al (2014) catalogued several classes of non-coding transcripts including over 70 eRNAs [421]. Knockdown of an NF-κB regulated eRNA upstream of the IL1B gene in a monocytic cell line led to reduced expression of IL1B [421]. Functional effects for eRNAs have recently been demonstrated in neurons by Schaukowitch et al (2014), with knockdown of transcripts from one enhancer being shown to decrease downstream transcription of mRNA from the *Arc* (*Activity regulated cytoskeletal protein*) gene, which is functional during activation of neuronal synapses by neurotransmitters [427]. While these reports of eRNA function at the transcript level are encouraging, the effects of these knockdown experiments are often small, both in terms of target knockdown and effects on downstream gene expression, and it is mechanistically unclear why such experiments should have an effect given that eRNAs are restricted to the nucleus and are already unstable.

1.3.3 Mechanisms of RNA function

Mechanisms for possible ncRNA functions are still uncertain, but their effects could conceivably include acting as decoys, scaffolds or guides for regulatory complexes [430] and additionally there is some evidence they

may influence the activity of such complexes through allosteric effects [431]. As guides, ncRNAs may form binding sites to recruit proteins such as chromatin-modifying complexes to specific genomic regions. As scaffolds, binding of an ncRNA to multiple complexes may allow orchestration and co-localisation of the independent activities of such complexes. As decoys, ncRNAs may sequester complexes away from alternative targets such as regulatory DNA sequences. The mechanistic and functional consequences of ncRNA transcription are in part influenced by their genomic origin but also by their structure and cellular localisation. Effects on gene expression and cellular identity can be mediated through both *cis*- and *trans*-regulatory mechanisms. This distinction between *cis* and *trans* regulation is related to the proximity of a non-coding locus to a gene with *cis*-acting ncRNAs being transcribed from the same locus as their target while *trans*-acting RNAs may have effects at a distant site with this interaction possible being mediated by looping of DNA to bring such sites into close proximity [432, 433] or by RNA diffusion.

1.3.3.1 Cis-regulation of gene expression

Promoter-associated non-coding transcripts (Section 1.3.2.3) are likely to act through *cis*-regulatory mechanisms and include the short ncRNAs described by our lab [416] and others [412, 413, 415] and also promoter upstream transcripts (PROMPTs) [417, 418, 434]. Those described by our lab, which are enriched at repressed polycomb target genes, precipitate with the PRC2 polycomb complex [416] and are predicted to form stem-loop secondary structures. The formation of secondary structure of ncRNAs has frequently been suggested to provide potential for specific protein-RNA interactions [435]. These structures may explain the targeted

recruitment of PRC2 to specific genes. Supporting this theory, many chromatin-modifying complexes are known to contain subunits capable of binding RNA *in vitro* [436, 437].

While direct evidence of functionality is still missing, the transcription of PROMPTs has been suggested to have an activatory effect on counterpart gene expression. It is been postulated that the presence of polyA sites upstream of TSSs limits transcription to the production of short, unstable transcripts upon bidirectional transcription at these promoters. This transcription may maintain the promoter in an accessible conformation, which is supported by an observation that transcription in the PROMPT window (0.5kb-2.5kb) upstream of the *Sphk1* gene is anti-correlated with repressive CpG methylation [438]. PROMPTs are preferentially degraded by the exosome, compared to their gene counterparts [417, 434] and this degradation is thought to enforce directionality of transcription and expression from the gene *in cis* [418]. In addition, PROMPTs may provide reservoirs of RNAPII activity in close proximity, ready to be harnessed for gene transcription.

Interactions with chromatin-modifying complexes have been suggested as possible mechanisms for several classes of ncRNAs. For example lincRNAs have also been reported to associate with histone modifiers and the chromatin insulator CTCF [439, 440]. The chromatin modifying complexes PRC2 and CoREST coimmunoprecipitate in native RNA immunoprecipitation assays with several lincRNAs, which has been suggested to support the existence of a physical association [441, 442].

Such associations have led to suggestions that an interaction between XIST and the repressive complex PRC2 is the basis for X chromosome inactivation [398]. In mammals, X chromosome inactivation in the cells of the homeogametic sex is required for dosage compensation. *XIST* RNA is upregulated on the future inactive X with a 5' internal transcript of *XIST*, repA (repeat A), reported to recruit PRC2 through the formation of stem-loop structures, leading to silencing of the chromosome [398]. The basis for this interpretation is that native RNA immunoprecipitation and electrophoretic mobility shift assays (EMSAs) show an *in vitro* association between repA and PRC2 and that in some circumstances knockdown of repA containing transcripts leads to a decrease in H3K27me3, which is deposited by PRC2 [398]. The extent to which interactions between ncRNAs and chromatin modifying complexes is specific is however controversial [443]. The *in vitro* assays, such as EMSAs, upon which these interactions are based may not effectively discriminate between specific and non-specific interactions but may rather reflect general structural properties of RNAs. Lack of crosslinking, in order to prevent RNA degradation, during native RNA immunoprecipitation may lead to non-specific *in vitro* associations that do not reflect the *in vivo* state. In addition the use of isotype controls reflects assay specificity rather than that of the protein itself. The more recently developed PAR-CLIP and iCLIP techniques combine UV crosslinking of RNA to protein with radiolabelling [444] and RNase protection assays and sequencing to better characterise interactions. Through such techniques, support for interactions between RNA and the PRC2 component JARID2 has been found [445]. Despite controversy surrounding the nature of its interaction with PRC2, support for

a *cis*-regulatory role for XIST itself in X chromosome inactivation is strong and is demonstrated by ectopic expression of an inducible *XIST* transgene leading to chromosomal condensation, even when inserted into autosomes [446].

While the inactive X condenses, expression of *XIST* from the future active X is downregulated in *cis* by its natural antisense transcript *TSIX*. There is some evidence that this downregulation may occur through recruitment of the DNA methyltransferase DNMT3a by *TSIX* leading to DNA methylation and repression of the *XIST* promoter [447]. Another natural antisense transcript, the *Zeb2* NAT, has been shown to have cis-regulatory effects with consequences for epithelial to mesechymal (EMT) transition [448], a hallmark of metastatic initiation for cancer progression. Zeb2 is a transcriptional repressor of E-cadherin, which is upregulated during EMT. Interestingly *Zeb2* mRNA positively correlates with its target, E-cadherin. This paradox is explained by an observed reduction in *Zeb2* translation, which is determined by alternative splicing of *Zeb2* leading to the exclusion of an internal ribosome entry site (IRES) sequence. This alternative splicing is dependent upon the antisense NAT [448].

LincRNAs with a positive effect on gene expression have been termed in some literature as ‘activating ncRNAs’ (ncRNA-a) [449], some of which have been shown to operate in *cis*. This is exemplified by ncRNA-7 (now annotated by Gencode as TRERNA1), which is required for activation of SNAI1 (Snail family zinc finger 1), a regulator of EMT. Depletion of ncRNA-7 using siRNA has been shown in functional assays to mirror the migratory

effects of loss of SNAI1 [449]. This RNA was shown to interact with components of the Mediator complex, a large multisubunit complex that associates with transcriptional machinery to regulate gene expression [450]. This interaction influences localisation and activity of Mediator, resulting in the transcriptional activation of SNAI1 [426]. Chromatin capture demonstrated an association between ncRNA-7 and the SNAI1 promoter, showing a physical interaction of these two loci despite their separation by kilobases of DNA. This interaction is explained by DNA looping, which allows the long range transcriptional activation effects of this ncRNA-a through a *cis*-regulatory mechanism [426]. Another ncRNA-a with *cis*-regulatory effects is HOTTIP. RNA immunoprecipitation and GST pull down indicates that HOTTIP may interact with the WDR5 subunit of the H3K4me3 histone methyltransferase complex MLL and has been hypothesised to recruit this complex to the HOX genes locus [451].

Cis-regulatory mechanisms suggested for eRNAs include increased RNAPII occupancy at the downstream target gene, as demonstrated by Mousavi et al (2013) [428]. In these experiments, siRNA was targeted against eRNAs transcribed in muscle cells and though RNAPII occupancy at the enhancer locus was not affected, it was reduced at the downstream gene [428]. The *Arc* eRNA described in neurons by Schaukowitch et al (2014) has been shown to support upregulation of the downstream *Arc* gene by acting as a decoy for the repressive complex negative elongation factor (NELF) [427]. This was demonstrated by knockdown of eRNAs using antisense locked nucleic acids (LNAs) (Section 5.2.8), which led to retention of NELF at the downstream gene in association with a reduction

in elongation RNAPII and mRNA synthesis.

1.3.3.2 Trans-regulation of gene expression

Trans-regulatory mechanisms are relevant to cytoplasmic ncRNAs but have also been demonstrated for nuclear localised ncRNAs with roles in transcriptional regulation.

The overexpression of an eRNA transcribed from a myogenic regulatory region leads to activation of expression of MYOG, which encodes a regulatory factor for myogenesis [428]. This effect observed by ncRNA overexpression demonstrates the functional activity of the ncRNA is independent of its genomic location and the act of transcription, while the association of this transcript with increased genomic accessibility suggests its activity is located within the nucleus [428]. In another example, transcription of the long ncRNA *HOTAIR* is thought to orchestrate *HOX* loci expression in *trans* by acting as a guide. The coordinated temporal and spatial expression of the *HOX* loci is fundamental to metazoan body plan organisation [452], with silent regions being repressed through PRC2 [453]. HOTAIR (HOX antisense intergenic RNA) is transcribed antisense to the *HOXC* chromosomal locus and has been suggested on the basis of native RNA immunoprecipitation to interact with PRC2 to target it in *trans* towards the *HOXD* locus. Consistent with this, siRNA depletion of HOTAIR leads to loss of H3K27me3 and increased transcription at the *HOXD* locus [454]. In addition, support for an interaction with the LSD1-CoREST histone demethylation complex, which removes H3K4me3, has been interpreted as evidence for a scaffolding role for HOTAIR in linking LSD1-CoREST and PRC2, allowing coordination of repressive chromatin

features [455]. The upregulation of *HOTAIR* has clinical relevance through an association with metastatic breast cancer that has prognostic value [456]. MALAT1, (metastasis associated lung adenocarcinoma transcript 1, also called NEAT2 (non-coding nuclear-enriched abundant transcript 2) has been suggested to perform a scaffolding role through an interaction with splicing factors and, through this, influences their localisation to nuclear paraspeckles with effects on alternative splicing [457]. Again, this assertion is controversial because subsequent work with a MALAT1 knockout mouse has revealed no obvious phenotype [458]. The lincRNA GAS5 (growth arrest specific 5) has a stem-loop structure that resembles the glucocorticoid response element upstream of glucocorticoid responsive genes. Through this mimicry, GAS5 acts as a decoy by interacting with the DNA binding site of the glucocorticoid receptor and in line with a functional role in sequestering the protein, GAS5 is upregulated in growth arrested cells [459]. Another example of a decoy RNA is PANDA (p21 associated ncRNA DNA damage activated), which binds the pro-apoptotic NF-YA transcription factor thus titrating it away from target genes [460]. The potential for allosteric influence of RNA on enzyme catalysis is demonstrated by the inhibition of CBP and P300 acetyltransferase activity by the TLS (translocated in liposoma) protein being promoted by binding of RNA to TLS [431].

1.3.4 Non-coding RNAs in T cells

Several ncRNAs have been identified as having roles in T cells, including during lineage specification. Recently, a profile of lincRNA expression across T cell differentiation in mice was performed identifying LincR-Ccr2-5'AS as being upregulated in Th2 cells and regulated by Stat6 and Gata3

[461]. Depletion of LincR-Ccr2-5'AS using shRNA resulted in reduced migration of Th2 cells to the lungs *in vivo* as well as a decrease in Th2 upregulated genes, although this effect was demonstrated as a genome-wide average rather than at specific genes. These data therefore suggest that LincR-Ccr2-5'AS acts as an activating lncRNA (ncRNA-a) [461]. Another lincRNA, NRON (noncoding repressor of NFAT) has been shown to augment the activity of the T cell transcription factor NFAT, as demonstrated through luciferase assays. NRON associates with several proteins in the cytoplasm including members of the importin-beta family, which mediate cytoplasmic to nuclear transport. Through this association, NRON regulates the cellular localisation of NFAT thus repressing its transcriptional activity [462]. A lncRNA downstream and antisense to the Th1 cytokine gene *Ifng* has been shown to be Th1-specific and positively correlate with *Ifng* transcription, although its expression is decreased rather than increased upon stimulation [463]. *Tmevpg1* (*Theiler's murine encephalitis virus persistence gene 1,* also called NeST: nettoie Salmonella pas Theiler's [in French] or cleanup Salmonella not Theiler's [in English]) was first characterised through the influence of its polymorphisms on susceptibility of mice to Theiler's virus [463] but its expression also leads to decreased Salmonella pathogenesis through increased Ifng production in CD8 T cells [436]. It has been suggested to act as an enhancer-like lncRNA (analogous to an ncRNA-a) and in humans (called IFNG-AS1), its position relative to *IFNG* is conserved and its expression is regulated by Stat4 and T-bet [464, 465]. Tmevpg1 is localised to the nucleus [436] and evidence supports a physical association of the RNA with the histone methyltransferase WDR5 and a correlation is seen

between Tmevpg1 expression and H3K4me3 at *Ifng* [436]. Ectopic expression of Tmevpg1 also increases Ifng expression showing that regulation is independent of genomic location and therefore occurs *in trans* [464].

1.4 The scope of this thesis

Through the experiments I present here, I have investigated the transcriptional regulation of T cell lineages, beginning with differential gene expression analysis to identify lineage-specific transcription. I start by focusing on Treg specific transcription (Chapter 3) and identify a set of ncRNAs that are preferentially expressed in Treg compared to Tresp and may exert control over Treg phenotype and function. I also investigate the dysregulation of these in the systemic autoimmune disease SLE, for which there is some evidence of Treg defects, particularly when naïve and memory subsets have been investigated. I have explored the transcriptional regulation of the Treg state further by identifying differences in transcription factor expression and histone modifications that relate to differences observed in SLE patents. In addition, I have explored the role of T-bet in Th1-specific transcription by investigating regions of high density binding, super-enhancers, through their production of eRNAs (Chapter 4). I have interrogated the functional roles of these Treg lincRNAs on Treg biology and of T-bet super-enhancer eRNAs on downstream gene expression using the *IFNG* gene and its counterpart super-enhancer as a model (Chapter 5). Finally, I present an overall discussion of this work in Chapter 6.

2 Materials and Methods

2.1 Solutions and buffers

Table 2.1: Solutions and buffers

Solution or buffer	*Constituent*	*Final concentration*
Complete media	Roswell Park Memorial Institute-1640 (RPMI, Sigma-Aldrich)	
	Foetal Calf Serum (FCS, Invitrogen)	10%
	Penicillin-streptomycin (Invitrogen)	100U/ml
Freezing media	FCS	
	Dimethylsulphoxide (DMSO, Sigma-Aldrich)	10%
Magnetic cell sorting (MACS) buffer	PBS (Invitrogen)	1x
	FCS	0.5%
	Ethylenediaminetetraacetic acid (EDTA, Sigma-Aldrich)	2mM
Xylene cyanol loading dye	Bromophenol blue	0.2%
	Xylene Cyanol	0.2%
	Glycerol (Sigma)	30%
	H2O	
Tris acetic acid EDTA (TAE) buffer	Tris (Sigma)	40mM
	Glacial acetic acid (Sigma)	0.1%
	EDTA, pH8 (Sigma)	1mM

2.2 Cell methods

2.2.1 Sample collection and PBMC isolation

Healthy donors and SLE patients were recruited in accordance with ethical approval obtained from University College London (ethics approval number 02/0241). Fifty millilitres of heparinised blood were collected from SLE patients (University College Hospital, London, UCLH) or healthy controls using butterfly needles (Therumo) into 50ml syringes (BD Biosciences). Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Hypaque (GE Healthcare) density centrifugation as follows. Blood was diluted (25ml blood to 15ml RPMI/5% FCS) in 50ml Falcon tubes (BD Biosciences) and 10ml Ficoll slowly layered underneath using a 10ml stripette (Fisher Scientific). Layered blood was centrifuged at 800g for 30min at 21°C with the break off during both acceleration and deceleration. The PBMC layer (situated between the Ficoll and plasma layers after centrifugation) was extracted using a pastette (Sarstedt), with care not to

disturb other layers, and deposited into a new 50ml Falcon tube containing 10ml RPMI (5% FCS). The isolated PBMC suspension was washed twice with RPMI (5% FCS) with centrifugation at 450g for 10min at 4°C.

2.2.2 Freezing of PBMC

Isolated PBMCs were resuspended in 10ml RPMI (5% FCS) and counted with a haemocytometer (Hycor). Cells were pelleted at 450g for 10min at 4°C and resuspended in freezing media (4°C) at $10x10^6$ cells/ml. Cell suspension was aliquoted into cyrovials (Nunc) at 1.5ml per 2ml vial and gradually cooled to -80°C in an isopropanol-containing cryo-container (Nalgene) before being transferred to liquid nitrogen.

2.2.3 Thawing of frozen cells

When required, cells were rapidly thawed by incubation in a water bath at 37°C until liquid and resuspended in 15ml of complete media (37°C). Cells were washed twice in complete media with pelleting at 450g for 10min at 4°C.

2.2.4 Flow cytometry staining and analysis

For intracellular cytokine staining, cells were incubated for 5 hr at 37°C in complete media with stimulation provided by phorbol myristate acetate (PMA, 40ng/ml, eBioscience) and ionomycin (2µg/ml, Sigma) and inhibition of Golgi export of protein provided by addition of monensin (1 in 1000, BD Biosciences). Cells for staining were washed twice with PBS and pelleted at 450g for 10min at 4°C. For the exclusion of dead cells an intracellular membrane amine reactive viability stain was used (Live/Dead ® Fixable Dead Cell Stain Kit, Life Invitrogen) as per the manufacturers instructions. Flow cytometry antibodies to extracellular antigens were added at their

titrated concentration in MACS buffer and incubated in the dark at 4°C for 20min. For staining of less than $1x10^6$ cells this was performed in 20µl in a 96-well U-bottom plate. For staining of greater numbers of cells (eg. for sorting) staining volumes were scaled up to 100µl per $10x10^6$ cells. Surplus antibodies were removed by washing in MACS with cell pelleting at 450g for 10min at 4°C. Cells were fixed with FoxP3 fix buffer (eBioscience) for 20min at 4°C. For detection of intracellular antigens, cells were washed in FOXP3 permeabilisation buffer (eBioscience), pelleted at 450g for 10min at 4°C and antibodies for intracellular antigens added at their titrated concentrations. Cells were incubated in the dark at room temperature for 1hr and surplus antibodies removed by washing in FOXP3 permeabilisation buffer. Cells were pelleted at 450g for 10min at 4°C and resuspended in MACS buffer. Flow cytometry acquisition was performed on a LSR Fortessa (BD Biosciences) or Image Stream (Amnis) and analysed on Flowjo software (Tree Star). Compensation was performed using anti-mouse Igk and negative compensation particles (BD Biosciences). Gating strategies and antibodies are detailed in Section 2.2.6.

2.2.5 Fluorescence associated cell sorting (FACS)

Cells were washed once with MACS buffer and pelleted at 450g for 10min at 4°C and extracellular antigen staining performed as per section 2.2.4. Cells were pelleted at 450g for 10min at 4°C and resuspended in MACS buffer at $40x10^6$ cells/ml and passed through a 35µm filter to remove aggregates. For the exclusion of dead cells, 1µl of DAPI (1mg/ml) was added 10 min prior to sorting. Cells were sorted on a FACS Aria II (BD Biosciences) and isolated populations collected in MACS (50% FCS).

2.2.6 Flow cytometry gating strategies and antibodies

Figure 2.1 illustrates gating strategies for flow cytometry analysis and cell sorting and Table 2.2 gives details of antibodies.

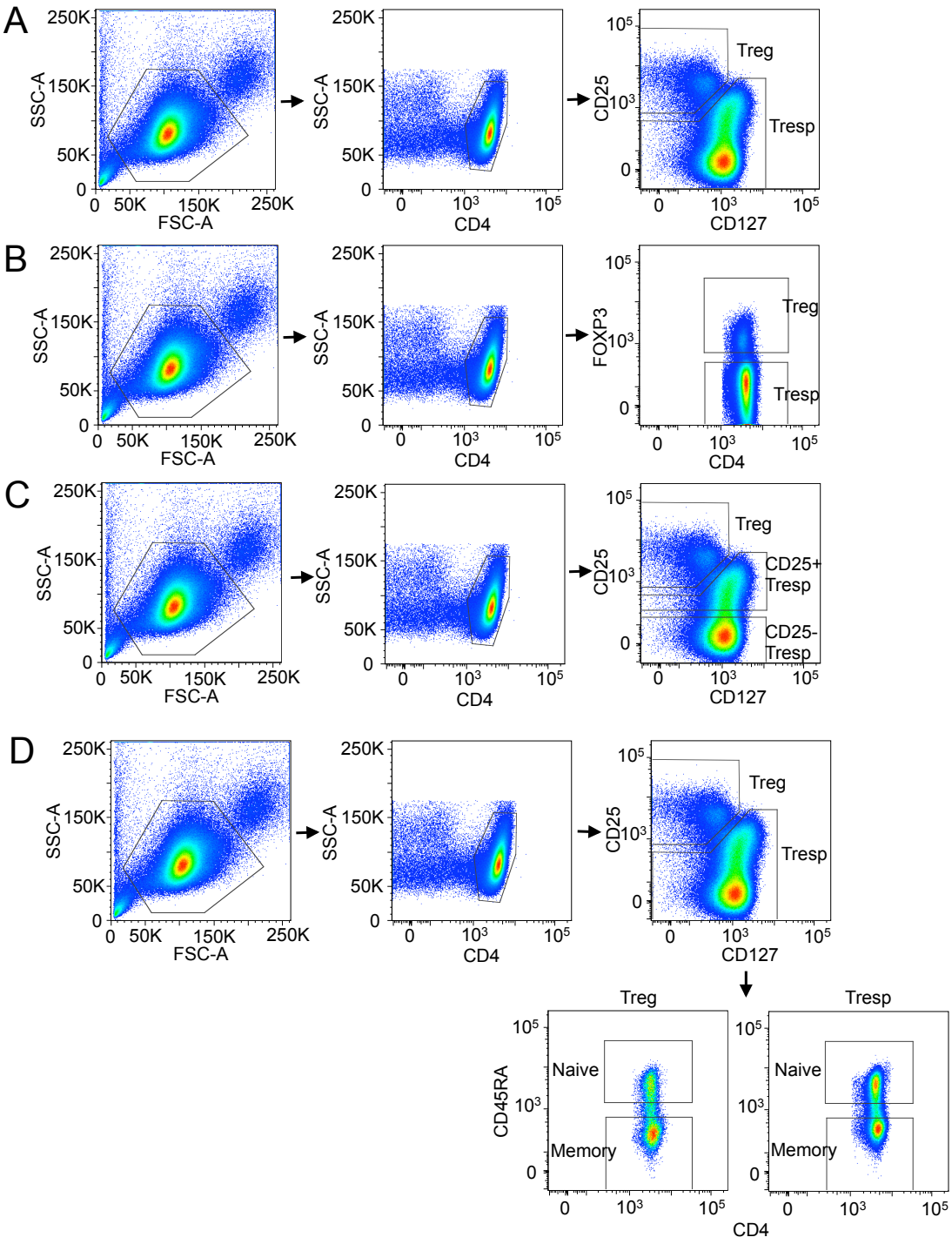


Figure 2.1: Gating strategies for flow cytometry analysis and cell sorting
Gating strategies for T cell subsets. A: Total Treg and total Tresp based on surface antigens. B. Total Treg and total Tresp based on the Treg lineage-specifying transcription factor FOXP3. C. Total Treg and CD25+ Tresp and CD25- Tresp. D. Naïve and memory subsets of Treg based on CD45RA expression.

Table 2.2: Flow cytometry antibodies

Antibodies were diluted to their titrated concentration in MACS buffer for fluorescence activated cell sorting or in FACS buffer for *ex vivo* and *in vitro* flow cytometry staining.

Corresponding plot in Figure 2.1	***Antigen***	***Fluorophore***	***Clone***	***Isotype***	***Concentration***	***Manufacturer***	***Catalogue number***
A.	CD4	AF700	RPA-T4	Ms IgG1, κ	1/80	BDBioscience	557922
	CD25	PE	M-A251	Ms IgG1, κ	1/5	BDBioscience	555432
	CD127	PerCP.Cy5.5	eBioRDR5	Ms IgG1, κ	1/10	eBioscience	45-1278-42
B.	CD4	AF700	RPA-T4	Ms IgG1, κ	1/80	BDBioscience	557922
	CD25	PE	M-A251	Ms IgG1, κ	1/5	BDBioscience	555432
	CD127	PerCP.Cy5.5	eBioRDR5	Ms IgG1, κ	1/10	eBioscience	45-1278-42
	CD45RA	FITC	HI100	Mouse IgG2b, k	1/40	BDBioscience	555488
C.	CD4	AF700	RPA-T4	Ms IgG1, κ	1/80	BDBioscience	557922
	CD25	PE	M-A251	Ms IgG1, κ	1/5	BDBioscience	555432
	CD127	PerCP.Cy5.5	eBioRDR5	Ms IgG1, κ	1/10	eBioscience	45-1278-42
D.	CD4	AF700	RPA-T4	Ms IgG1, κ	1/80	BDBioscience	557922
	FoxP3	AF647	236A/E7	Ms IgG1	1/50	eBioscience	51-4777-42
	Used alone and also in combination with antibodies described in A, B and C.						

2.2.7 In vitro Treg induction

TCDD (2, 3, 7, 8-Tetrachlorodibenzo-ρ-diooxin) inductions were performed as described by Gandhi et al. (2010) [88]. Naïve CD4 T lymphocytes were isolated as described in Section 2.2.5 and cultured in 96 well flat-bottom plates at $300x10^3$ cells per well in 200µl complete media with plate bound anti-CD3 (1µg/ml, UCHT1, eBioscience), soluble anti-CD28 (1µg/ml, CD28.2, eBioscience), IL-2 (50U/ml, Peprotech), human recombinant TGF-β (10ng/ml, R&D) and TCDD (100nM, Cambridge Isotype Laboratories). After 6 days, cells were analysed by flow cytometry as described in Section 2.2.4.

2.2.8 In vitro Treg expansion

Treg were isolated as described (Section 2.2.5) and cultured in a 96-well flat-bottom plate at $2x10^5$ per well with 100nM rapamycin, 1000IU/ml IL-2 and at a 1:1 ratio with Human T cell activator CD3/28 Dynabeads ® (Life Technologies). IL-2 was replenished at day 2 and cultures split 1in2 whenever media appeared visibly exhausted. Cells were harvested after 8 days.

2.2.9 In vitro Th1 polarisation

Naïve cells were obtained as per Sections 2.2.1, 2.2.5 and 2.2.6. Naïve CD4 T cells exit the thymus expressing long 'RA' isoform of the cell surface receptor, CD45. After TCR stimulation, the gene encoding the receptor, PTPRC, is alternatively spliced eventually leading to expression of the 'RO' isoform on memory T cells. Repeated restimulation *in vitro* has shown re-expression of CD45RA can occur in non-naïve cells [466]. Therefore additional markers are required to isolate naïve T-lymphocytes.

This includes a lack of CD25 expression, which is upregulated after antigen encounter, and the chemokine receptor CCR7, hence naïve CD4 T cells can be identified as CD45RA+CD45RO-CD25-CCR7+. Cells were cultured at 37°C, 5% CO2 in complete media supplemented with IL-12 (10ng/ml), anti-IL-4 (10µg/ml) and IL-2 (100IU/ml). Stimulation was provided with CD3/28 activation beads at a 1:1 ratio with cells. At day 2 or 3 (whenever media was visibly exhausted) media was replaced. At further time points cells were split 1 in 2 as needed. Cultures were harvested at various time points up to 13 days. For cultures longer than 7 days, activation beads were replaced to provide an additional round of stimulation. Cells were harvested as 'unstimulated' or restimulated with either PMA (40ng/ml) and ionomycin (2µg/ml) or CD3/28 activation beads (1:1 ratio with cells).

2.2.10 Transfection for siRNA-mediated knockdown

siRNA were designed using Dharmacon's proprietary online siRNA design algorithm [467]. siSTABLE modifications (proprietary knowledge) were added to enhance efficiency of siRNA knockdown, a requirement for T cells [468]. siRNA were resuspended in PBS as per the manufacturer's instructions. A FITC-tagged non-targeting transfection indicator (siGLO ™, Dharmacon) was included at a 1:1 ratio with the targeting siRNA to assess siRNA uptake by flow cytometry at 24hr post transfection. Knockdown was assessed relative to a non-targeting control. siRNA sequences are detailed in Table 2.3. In order to transfect siRNA into cells, two electroporation devices were used (Sections 2.2.10.1 and 2.2.10.2).

Table 2.3: siRNA sequences

Target	*ID*	*Sense (5' to 3')*	*Antisense (5' to 3')*
FOXP3		TCGAAGAGCCAGAGGACTT	AAGTCCTCTGGCTCTTCGA
	SMARTpool (pool of 4)	Proprietary knowledge, Dharmacon.	Proprietary knowledge, Dharmacon.
ncRNA30	1	CCTCTGAAATTGTGAGACA	TGTCTCACAATTTCAGAGG
	2	GATGGGAATTGTTGGATAA	TTATCCAACAATTCCCATC
miR-146a		GGTGAAATCTCAACCGACA	TGTCGGTTGAGATTTCACC

2.2.10.1 Nucleofector ™ 2b Device, Lonza

Cells were washed in PBS and supernatant removed. Cells were resuspended in 100µl room temperature Nucleofection solution ® (Human T cell Nucleofection kit, VPA-1002, Lonza) with the addition of siRNA and transferred to an electroporation cuvette. For *ex vivo* unstimulated cells, $5x10^6$, or as many as possible, were used per condition and electroporation program U-014 used, as per the manufacturer's instruction [469]. This program was recommended for increased viability. For expanded, stimulated T cells, $1\text{-}5x10^6$ cells were used per condition and electroporation program T-023 or T-020 used, as per the manufacturer's instructions [469]. Complete media (500µl, 37°C) was added and cells transferred to a well of a 12 well plate containing 1.5ml 37°C complete media. Media was replaced 5hr post-transfection.

2.2.10.2 Agile Pulse Electroporation System, BTX

Cells were activated through a protocol optimised by Laurie Menger (Sergio Quezada, Karl Peggs) as follows: cells were cultured in a 12 well plate at $1x10^6$ per ml with activation beads (1:1 ratio with cells) in X-Vivo™ 15 media (Lonza), 10% FCS and 300IU/ml IL-2. After 3 days cells were washed in cytoporation media (BTX) and resuspended at $5x10^6$ per 200µl for each condition with addition of siRNA for electroporation.

2.2.11 Culture for Locked Nucleic Acid (LNA)-mediated knockdown

LNA were designed using Exiqon's proprietary online antisense inhibition design algorithm [470]. LNA were resuspended in H2O as per manufacturers instructions. A FAM-tagged non-targeting control was included at a 1:1 ratio with the targeting LNA to assess uptake by flow cytometry. Knockdown was assessed relative to the non-targeting control. LNA sequences are detailed in Table 2.4. Cells were activated with aCD3/28 beads (1:1 ratio) in complete media with the addition of LNA the media at various concentrations and cells harvested at the required time point.

Table 2.4: LNA sequences

Target	*ID*	*Sense (5' to 3')*	*Antisense (5' to 3')*
GAPDH		TTGTCATACCAGGAA	TTCCTGGTATGACAA
MALAT1		GGCATATGCAGATAATGTTC	GAACATTATCTGCATATGCC
CRNDE	1	CAATCCAATAAAGAC	GTCTTTATTGGATTG
	2	AAGTCAACCTATAC	GTATAGGTTGACTT
	3	AAACTGGCAATCAA	TTGATTGCCAGTTT
	4	AATAGCCAGTACAG	CTGTACTGGCTATT
	5	GCCGCAAGAGGCGG	CCGCCTCTTGCGGC
TBX21 eRNA 2 plus	1	CCTGCTAATCGAAC	GTTCGATTAGCAGG
	2	AGAGGTGACACAATT	AATTGTGTCACCTCT
IFNG eRNA 3 minus	1	CCACAAGCCATTTTAA	TTAAAATGGCTTGTGG
	2	TACTTCTGTATGCTGA	TCAGCATACAGAAGTA
IFNG eRNA 1 minus	1	ATAGTAAGACTGCAAG	CTTGCAGTCTTACTAT
	2	CAGGGATTTGGGTCAG	CTGACCCAAATCCCTG
IFNG eRNA 2 plus	1	ACATGAGGATAACTTG	CAAGTTATCCTCATGT
	2	TTGACAGCTCACTTCT	AGAAGTGAGCTGTCAA
IFNG eRNA 7 plus	1	TGAGTGCCTTTAAGAA	TTCTTAAAGGCACTCA
	2	GTTACAGGTTGGAGAA	TTCTCCAACCTGTAAC

2.3 RNA methods

2.3.1 RNA extraction

Cells intended for RNA extraction were pelleted at 450g for 10min, supernatant removed and cells resuspended in 1ml Trizol® (Invitrogen) for up to $10x10^6$. Alternatively, for small cell numbers, cells were resuspended in 250l Trizol LS ® (Invitrogen) and cells and stored at -80°C until extraction. Upon thawing, 200µl of chloroform was added per 1ml Trizol®

and samples shaken by hand for 15s. Phases were separated by centrifugation at 12,000g for 15min at 4°C. The aqueous phase, containing RNA, was transferred to a new RNase-free 1.5ml tube (Eppendorf). Organic phases, containing DNA and protein, were returned to -80°C. Isopropanol (500µl, Sigma) and 20µg of glycogen (Roche) were added to the aqueous phase and incubated at room temperature for 10min. Precipitated RNA was pelleted by centrifugation at 10,000g for 10min at 4°C. Supernatant was discarded and pellet washed with 1ml room temperature 75% ethanol (Sigma). RNA was pelleted again with centrifugation at 6000g for 5min at 4°C, supernatant discarded and pellet air-dried at room temperature for 5min. RNA pellets were resuspended in 25µl RNase-free H_2O and concentration determined by spectrophotometry (Nanodrop 2000, Thermo Scientific).

2.3.2 DNaseI treatment of RNA

DNaseI reactions were performed in 50µl with 10x DNase TURBO™ buffer (Ambion) and 50x DNaseI TURBO™ enzyme (Ambion) with incubation at 37°C for 30min. DNase inactivation reagent (5µl, Ambion) was added, incubated for 2min at room temperature and inactivation reagent pelleted by centrifugation at 10,000g for 1.5min. Supernatant (45µl) was transferred to a new tube and 150µl of 100% ethanol, 5µl sodium acetate (Sigma) and 20µg glycogen added. RNA was precipitated at -20°C for 1hr and pelleted by centrifugation at 18,000g for 30min at 4°C. Supernatant was discarded and RNA washed with 1ml 80% ethanol (-20°C) and pelleted at 10,000g for 10min at 4°C and supernatant discarded. RNA pellets were air-dried at room temperature for 5 min and resuspended in 25µl RNase-free ddH_2O and concentration determined by spectrophotometry (Nanodrop 2000,

Thermo Scientific).

2.3.3 Assessment of RNA quality

Extracted and DNaseI treated RNA was assessed for the presence of contaminants by spectrophotometry (Nanodrop 2000, Thermo Scientific). Absorbance ratios of 260nm to 280nm or 230nm of greater than 1.8 were presumed to be indicative of high purity. RNA degradation was assessed on an Agilent 2100 Bioanalyzer using an Agilent RNA 6000 Picochip according to the manufacturers instructions [471]. An RNA Integrity Number (RIN) greater than 7, or comparable electropherogram, was considered to indicate sufficient quality for microarray or RNA-Seq gene expression analysis.

2.3.4 Reverse transcription

Up to 5µg of DNaseI treated RNA was made up to a volume of 11µl with H_2O and were incubated with 1µl primers (either oligo$(dT)_{20}$, 50µM, Invitrogen P/N58063, or random hexamers, 0.2µg/µl, Thermo Scientific, ♯50142) and 1µl dNTP mix (10mM each dATP, dGTP, dCTP, dTTP, Promega) at 65°C for 5min followed by 1min on ice to anneal primers. Four microliters of First Strand Buffer (5x, Invitrogen), 1µl DTT (0.1M, Invitrogen), 1µl RNaseOUT™ Recombinant RNase Inhibitor (40 units/µl, Invitrogen) and 1µl reverse transcriptase (SuperScript® III, 200 units/µl, Invitrogen) were added (final volume 20µl). The reaction was incubated at for 10min at 25°C followed by 50min at 50°C and finally 15min at 85°C to inactivate the reverse transcriptase enzyme.

2.3.5 Labeling and amplification of RNA for microarrays

RNA was labelled with Agilent Low Input Quick Amp 2 Colour Labelling kits

according to the manufacturers instructions [472]. The reagents detailed below were provided with this kit unless otherwise stated. Sample RNA (i.e. RNA from SLE patients or healthy donors) was mixed with spike-in control (Agilent, # 5188-5279) A and labelled with Cy3-CTP. Reference RNA (Stratagene Universal Human Reference RNA, Agilent) was mixed with spike-in control B and labelled with Cy5-CTP. RNA template was annealed with T7 promoter primer and first strand cDNA synthesised with First Strand Buffer, 0.02M DTT, 1mM dNTP and Affinity Script RNase Block Mix with incubation at 40C for 2hr. Affinity Script enzyme was inactivated by incubation at 70C for 15min. Complementary RNA was transcribed with Transcription Buffer, 0.01M DTT, 1x NTP mix (which included fluorescently labelled CTP) and T7 RNA Polymerase Blend with incubation at 40C for 2hr. Labelled RNA was purified using the Qiagen RNeasy mini kit and spin columns according to manufacturers instructions [472]. Labelled RNA was hybridised at 60C for 17hr to Sureprint G3 Human Gene Expression microarrays with 8x60K format (Agilent #G4851A). Incorporated fluorescence was read on an Agilent High Resolution C scanner.

2.3.6 RNA-Seq library preparation

Prior to RNA-Seq library preparation all RNA was DNaseI treated (Section 2.3.2) and quality assessed (Section 2.3.4). A RIN of 7 or above, or comparable electropherogram was considered of sufficient quality for RNA-Seq. All libraries were prepared with 1µg of DNaseI treated total RNA as starting material. Preparation of RNA for polyadenylated RNA libraries required purification of polyadenylated RNA (Oligotex™, Qiagen) as per the manufacturer's instructions. Total RNA library preparation required

removal of ribosomal RNA (Ribo-zero™ Gold, epicentre) followed by removal of 5' cap structure (Section 2.3.6.1). All libraries required fragmentation and repair (Section 2.3.6.2) prior to library generation (NEBNext ® Multiplex Small RNA Library Prep Set for Illumina, NEB #E7300S/L) and gel purification (Section 2.3.6.3).

2.3.6.1 Removal of 5' cap structure

RNA, in 7µl, was incubated with tobacco acid pyrophosphatase (TAP) buffer (Ambion) and 2µl TAP enzyme (Ambion, 2380G) at 37 ^{o}C for 1 hour. To re-purify RNA the reaction was made up to 115µl with H_2O and 15µl ammonium acetate and 150µl acid phenol chloroform added. After vortexing and centrifugation at 13,000g for 5 mins, the aqueous phase was removed to a new eppendorf containing 150µl chloroform. Centrifugation was repeated and RNA from the aqueous phase was precipitated with 150µl isopropanol. After a 10-minute incubation at room temperature, RNA was pelleted by centrifugation and pellet washed with 80% ethanol, supernatant removed and pellet air dried for 5 min before being resuspended in H2O.

2.3.6.2 Fragmentation and repair

RNA was made up to a volume of 182µl with H2O and incubated with Tris (pH8, 40mM final concentration), magnesium acetate (30mM final concentration) and potassium acetate (100mM final concentration) in a total volume of 200µl at 94°C for 4min. Reactions were placed immediately on ice and 20µl 0.5M EDTA added. Fragmented RNA was precipitated with 20µl sodium acetate, 500µl ethanol and 20µg glycogen at -20°C for 1hr. RNA was pelleted by centrifugation at 13000g for 30min and pellets

washed with 500µl 75% ethanol (-20°C) with centrifugation at 13000g for 5min. Pellets were air-dried for 5min and resuspended in 16µl H2O. Two microliters of 10x phosphatase buffer (NEB), 1µl of Antarctic phosphatase (NEB) and 1µl RNaseOUT (Invitrogen) were added (final volume 20µl) and reactions incubated for at 37°C for 30min followed by 5min at 65°C. Reactions were made up to 50µl with 17µl DDW, 5µl 10x PNK buffer, 5µl ATP (10mM), 1µl RNaseOUT and 2µl PNK and incubated at 37°C for 1hr.

RNA was purified using RNeasy kit reagents (Qiagen) according to the following protocol: Fragmented RNA was made up to 100µl with H2O and 350µl RLT buffer added and sample vortexed. Six hundred and seventy five microliters of ethanol were added and the sample mixed by inverting the tube. Sample was applied to an RNeasy MinElute column in a collection tube, centrifuged and flow-through discarded. Five hundred microliters of RPE buffer was applied to the column, centrifuged and flow through discarded. The same process was performed with 750µl 80% ethanol. The column was then centrifuged with the lid open for 5min and residual liquid removed from the rim of the column with a pipette. The column was transferred to a clean 1.5ml eppendorf tube and 15µl H2O applied to the centre of the membrane and RNA eluted by centrifugation.

2.3.6.3 Gel purification of RNA-Seq library

Ten microliters of Xylene Cyanol loading dye were added to each PCR library before loading onto a 2% agarose TAE gel with SYBRsafe nucleic acid stain. Gels were electrophoresed at 100V for 40min or until primer dimer bands were resolved from library product bands. DNA was visualised on a dark reader and bands between 130bp and 350bp (relative to Hyperladder V, Bioline) excised with a sterile scalpel. DNA was

extracted from the gel using the Qiagen MinElute kit according to manufacturer's instructions but without heating gel in QG buffer and with an extra PE buffer wash. DNA was eluted in 15µl EB buffer, preheated to 50^{o}C.

2.3.6.4 Library sequencing

Libraries were quantitated by qPCR (Library quantification kit, KAPABIOSYSTEMS) and average fragment size analysed by Bioanalyzer (DNA HS assay, Agilent). Libraries were sequenced on an Illumina HiSeq with 50bp paired end sequencing.

2.4 DNA methods

2.4.1 Primer design

Primers were designed using the Roche Universal Probe Library Assay Design Centre [473, 474] or Primer3 [475]. Suggested primer pairs were tested by *in silico* PCR (UCSC genome browser) to ensure specificity for producing a single amplicon.

For QPCR validation of microarray data, primers were designed as close as possible to the microarray probe target sequence. Primers were synthesised (Sigma-Aldrich) and tested by QPCR (Section 2.4.2) on cDNA using a four-stage 1 in 5 serial dilution. Primer pairs with an efficiency rate of 90-105% and sequence specific amplification (melting curve analysis) were considered admissible. All primer sequences are given in Table 2.5.

Table 2.5: Primer sequences

Presence of a probe sequence indicates use of TaqMan™ chemistry. All other primers were used with SYBR® Green. Hs: Human; Ms: Mouse.

Target	*Species*	*ID*	*Forward (5' To 3')*	*Reverse (5' To 3')*	*Probe (5' to 3')*
ACTB	Hs		AGCACAGAGCCTCGCCTTT	TCATCCATGGTGAGCTGGC	
	Mm		TCTTTGCAGCTCCTTCGTTG	ACGATGGAGGGGAATACAG	
HPRT	Hs		AGTCTGGCTTATATCCAACACTTG	GACTTTGCTTTCCTTGGTCAGG	6FAM-TTTCACCAGCAAGCTTGCGACCTTGA -TAMRA
	Mm		TCAGTCAACGGGGGACATAAA	GGGGCTGTACTGCTTAACCAG	
GAPDH	Hs		GGCTGAGAACGGGAAGCTT	AGGGATCTCGCTCCTGGAA	6FAM-TCATCAATGGAAATCCCATCA-TAMRA
	Mm		ATGATGCGCAAAGGTATGCA	CCCCATCTCCCCCTTCCT	
ncRNA30	Hs	A	GGCTCCATGCCAGTCCTAT	TTTGTGAGCATTGGCACAG	
		B	ATGGCAGAAGGTCTAGTGGG	GGCACTTGAACCTCACAACA	
		C	CATTGTATGGAGGGGTTCTGC	ACTTGAACCTCACAACAGAACT	
		D	CAAATGGGATGCAGTGGACC	GGCACTTGAACCTCACAACA	
miR146a			ATGAAAGAGTAACCTGTCCCAGA	AGAATGAGCATGCTGTGGAGT	
CRNDE		A	CAGCCGTTGGTCTTTGAAAT	AAACCACTCGAGCACTTTGA	
		B	GGCGCTAACGGTCGGTAA	ACCAGCCTTGGGATGAATTT	
		C	TGAACTAAGGGGTTCCTCCA	TTTGGTACCCATGACTGGAA	
		D	GATCGCGCTATTGTCATGGAG	CTTCAGAATCTCCTCCTTCCA	
FOXP3	Hs		GAAACAGCACATTCCCAGAGTTC	ATGGCCCAGCGGATGAG	
IRX5			TGGCTAAAGACCCGAAAATG	GGGAGAGTCTTTGCACAGGT	
NR4A3			GTGTCTCAGTGTTGGAATGGT	TTTGGTTTGGAAGGCAGACG	
MALAT1			GACGGAGGTTGAGATGAAGC	ATTCGGGGCTCTGTAGTCCT	
MKI67			ACGAGACGCCTGGTTACTA	GCTCATCAATAACAGACC CATTTAC	
CB144112			AGTAACCTGTGATCTTCCCCT	GGGGTTGCCTACTCTTCTGT	
GPR15			TGTTGAGCATTGTGACCTGC	CCTGATTGCTGGTAATGGGC	
IFNG	Hs		AAACGAGATGACTTCGAAAAG	ACAGTTCAGCCATCACTTGG	
	Mm		GCCAAGTTTGAGGTGAGACG	GTGGACCACTCGGATGAGC	
IFNG eRNAs	Hs	A2	CAGAGAAATGCCAGCAAAAAC	ACCCTCGTTGTGCTAGTTCG	
		A3	CTGGAATGTCTCCACTGAGC	CTCAGGAAGGGTTTGTGAGC	
		B1	ACCTGAGCCAGCTCCTATCC	CGGTAGAGGGGAAAAACAGC	

Target	*Species*	*ID*	*Forward (5' To 3')*	*Reverse (5' To 3')*	*Probe (5' to 3')*
		B2	TCGCTTTCCTTTCTCCAGTT	TGACAAACCCTCTTACAACCA	
		C1	GGATGTGCTGTTTTTCACACC	CTGGAGCAATCCTATCACAGC	
		D1	AGCATGCTGGAAAATGAACC	AAGATGCTTGGGTGTGTGC	
		D2	TTCAAGGTCCCTGTTTCTGC	GCAGAAGATGCTTCCTCCAG	
		E1	GGCCAAGTCTTTCTCAGAGC	TTCCCTGGACTGTTGTAGGC	
		E2	GCCTGAAGTCCCTATGTTGC	TGATTTGCACACTGCTCACC	
		F1	GCCCACACTTGGTTTAATGC	GTGGAAACAAGTGCAAAACG	
		F2	CTTTGTTGGTTTGGCCAGTT	GTCACACCTTGGCTCCTGAT	
		G1	ATGTCCTTCCTGTGGGTCAC	CTGGGCTGAAGAGAGGAATG	
		G2	CCTCATGTGCTTTTCTTCCTC	ACTCAGCCAGGCACTTTGAG	
		G3	TAGTGCCCAAGCCACATAGG	TGAGATTGGCAGTAACACAGC	
		H1	GGAGCATGAACAGATGTGAGG	GTGTGCAGGGCTCCTATGC	
		H2	TCTAGCCACCCTAAGCATCC	GGTTTCCTCTCAGGGTACTGG	
		I1	ATGACCCCAAGACTGAAACG	AGGTGGGACTTCACCAAGG	
		I2	CAGGGGCACAATGAAAGC	GCTATTGGCATACAGCATGG	
		J2	TTGAATAAGCCGATTGTTTGC	GAGCATAGGACTCCCTGTGG	
		J3	CTCTTCAGGGGTGTTCATGG	ATGCAGGGTCTGATCTGAAGG	
		S	CAGTCACAGGCAGAGGGTAA	TACTACCTGCCTTGCCAACA	
		T	TGATACCCAACTAACATGCCAA	TCAGGAGTGCTTTTCAACTCT	
		R	CGAACATGCACAGTGACCTC	AGCAGGACCAGGATCCTTTC	
	Mm	Aii1	CTTCATGCTTGGTCCTTTGG	CAGGGCTACACAGGAAAACC	
		Aii2	TGCAGACAGGGAAATAAAAGC	AAGTCACCCCACCATAGAGC	
		Bi1	CCAAGATGAGTCAGCACAAGG	CACGCGCCAGAATAAATAAG	
		Bi2	TATTTGCCCGCTGTTGACC	GTAAGGCATGTTCTTGACTGC	
		Bii1	AAGGCCCTGACTTCATTAGC	GGCTCTCCTCTGCAGAATCC	
		Bii2	AATTGTGGTCGTTGTGTCTCC	GCCTGGGTTTCTGATACAGC	
		Ci1	GTCGCTAGGAGCACTATACCG	CTTGCAACTTAAAGCCAAAGC	
		Ci2	CAAATTAACCTCAGCCATGC	TGGCCATTTAAAGCTTCTCC	
		Cii1	CAACTGAGGCATTAGCAGAGC	TCGTACAGAAAGGACTTCACG	
		Cii2	CTCAGCTGCAAAAGAAAAGG	TGTAACCCTGCTTCTGACTCC	

Target	*Species*	*ID*	*Forward (5' To 3')*	*Reverse (5' To 3')*	*Probe (5' to 3')*
		Cii3	GTCGTTAACTTTCCCGATCC	AGCCCACACAGGAATCTGAG	
		Cii4	GATGGCTTGCTCACATACTCC	AACCATGGAAACTTGTCAGG	
		Cii5	TTAGCTTGCCCTGTTCTTGC	TGGATCGTGCAGCTAAGAGG	
		Cii6	AGCTTGACTTTCATGCTTCC	GCTACCCATTTCCCAGAGG	
		Ciii1	CGAAGGGGGAAAGATGTACC	AGCATCTCCTCCTTTTCTTCC	
		Ciii2	AGTGCCCACTGCAATACAGG	AAGGGCCTCAACTCTTTGC	
		Di1	CTTCCTTTCTAGGCCTGTCC	CATGCCATGTATCAGTGTTCC	
		Di2	GACAGAGGAAGAGGGTGTGG	AACTCAGGGGACTGGTTTCC	
		Ei1	TTTGATTGTTTGCACAAGACC	AGTTAGAAGTGGGGCTCTGG	
		Ei2	TCATGCTGTTGAAAGAAATGG	TTTGTGCTTTCCTGATTCACC	
		Fi1	TTCTGCAGGCTCACTATTGG	GTGCGCTGCCTGTAAAGC	
		Fi2	GCGCAACCAGGATGTAGG	CTGGTGGTGGATGTAACAGG	
		Gi1	TCCGTGTGACATGTCGTTTAG	AACAGAAGCCCTGCATTTTG	
		Gi2	CAGCTAAAGCCCCAAGAATG	TGGGAACTGGATTGTGTTTG	
		Hi1	TCTGGAGTAGGGATGGAAGG	GACCCAGATGACCAAATTAGC	
		Hi2	TCTTGTGCTTTCATCTTTTGC	CCCAAACTTGGTGCTACCTG	
		Hii1	AAAAGCCCAGAGTGTCAACC	GCTCTTCTCTTTCCAAGAAGC	
		Hii2	AGCTCCCATTAATGACACACC	GTGGTAACACACACACACACC	
		Ji1	CACACACACATGGATGCAC	AGCCAGAACAGACAGACAGAC	
		Ji2	CACTTGGGAGGTAGAGACAGG	TATACCCTTGGCTGTTCTGG	
		S	CAAGGGTTGAGAATGGGTGC	GCCATCACTCTCTTGGGGAT	
		S	TGAGGGCTAGATGGAATGGT	TTGAAGATCACTCCTGCAAGT	
		R	GTGCTGTGACTGTAAGGTGC	ACAGGCTCACTTTCCTCTCC	
TBX21 eRNAs	Hs	1	CGAAGGAAGCCGTTTAGAGC	AGAAACCGCACAGTGTTCAC	
	Mm		CTGAGACACGGCTCTGTCA	AATGGTCTGAGGGTGAGGC	
	Hs	2	TGTGTCACCTCTCTGAGCAC	AGAGGACACTATCGTTCCCA	
	Mm		AGGGTGGTAGTACAGGCTTG	CCATCCTCCAGCCTCAGG	

2.4.2 qPCR

Q-PCR was performed using QuantiTect SYBR green PCR kit (Qiagen) chemistry if possible or QuantiFast Probe RT-PCR Plus kit (Qiagen) alternatively. Reactions were performed in 10µl with 5ng cDNA template, 300nM primers and 250nM probe (when relevant). Cycling conditions were: 95°C (10min) followed by 40 cycles of 95°C (15s) and 60°C (1min). All reactions were performed in triplicate in an Applied Biosystems® 7500 Fast Real-time PCR system with no reverse transcriptase and no template (water) negative controls. If the standard deviation of number of cycles to threshold (ct) exceeded 0.5ct, outlying replicates were excluded.

2.5 Bioinformatics methods

2.5.1 Analysis of microarray data

RNA was labelled, amplified and hybridised to slides as described in Section 2.3.5. Microarray slides were scanned with an Agilent High Resolution C Scanner. Fluorescence intensities were converted into log base 10 expression ratio of Cy5 to Cy3 using Feature Extraction software (Agilent, v10.7). Data was extracted and converted into inverse log base 2 expression ratios of Cy3 (sample) to Cy5 (reference) signal in Excel (Microsoft). For normalisation, expression ratios were iteratively median centred ten times across both genes and arrays and filtered to remove genes with less than one value across all arrays (performed in Cluster 3.0 [476]).

2.5.2 Hierarchical clustering

Hierarchical clustering was performed using Cluster 3.0 [476] on median centred log2 expression ratios. Pairwise Pearson correlation coefficients

were calculated using Cluster 3.0 and unsupervised hierarchical clustering performed by average linkage analysis. Heatmaps and dendrograms were viewed with Java Treeview [477].

2.5.3 Principle component analysis

Log2 expression ratios of sample to reference were uploaded to Cluster 3.0 in a text file. Expression ratios were iteratively mean centred ten times across both genes and arrays (performed in Cluster 3.0). Eigenvectors of the covariance of the matrix of the data were calculated by Cluster 3.0 and principle components exported as a text file for analysis in Microsoft® Excel®.

2.5.4 Ranksum and rankproduct analysis

Ranksum and rankproduct analysis were performed using RankProdIt [478-481]. Log2 expression ratios of sample to reference (log base 2) were iteratively median centred ten times across both genes and arrays filtered to remove genes with less than one value across all arrays (performed in Cluster 3.0). Text files containing the data were uploaded to RankProdIt [481], allowing comparison of two conditions (cell subsets) at a time to identify significantly up and down regulated transcripts. To identify differential expression unique to each T cell population, the population in question was treated as ‘condition 1’ and all other populations treated as ‘condition 2’. To interrogate Treg or Tresp as a whole population, rather than individual naïve and memory subpopulations, naïve Treg and memory Treg were grouped as ‘condition 1’ and compared with naïve Tresp and memory Tresp as ‘condition 2’. RankProdIt calculates the ‘rank product’ for each gene being upregulated in an array and its combined probability of

upregulation over replicate arrays, as shown in Equation 2.1.

Equation 2.1: Calculation of rank product

RP: Rank product; g: gene; up: upregulated; k: number of replicates; i: replicate; r: rank; n: number of genes; Π = cumulative product.

$$RP_g^{up} = \prod_{i=1}^{i=k} (r_{i,g}^{up} / n_i)$$

Rank sums were calculated similarly (Equation 2.2):

Equation 2.2: Calculation of rank sum

RS: Rank sum; g: gene; up: upregulated; k: number of replicates; i: replicate; r: rank; n: number of genes; Σ = cumulative sum.

$$RP_g^{up} = \sum_{i=1}^{i=k} (r_{i,g}^{up} / n_i)$$

Rank products and sums for downregulated genes were similarly calculated. Genes were sorted by pfp value (proportion of false positives, analogous to a false discovery rate), a lower pfp value corresponding to greater statistical significance. A cut-off pfp value of 0.05 was taken.

2.5.5 Identification of non-coding transcripts

Non-coding transcripts were identified using publically available gene annotations accessed through the UCSC Genome Browser [482]. Evidence came from several international consortia. Details of their annotation process and curation are detailed in Table 2.6.

Table 2.6: Evidence for gene predictions and non-coding annotations

These tracks are publically available from the UCSC genome browser.

UCSC Genes
Gene predictions are based on RefSeq (see below), Genbank [483], CCDS [484], UniProt [485] and Rfam [486]. The Kozak consensus sequence (start-codon) and orthologous predicted proteins from other species are used as evidence of protein-coding capacity. It also considers nonsense-mediated-decay and start-codons in any frame upstream of the predicted start codon as evidence of lack of coding potential. [487].
ENCODE/GENCODE v12
The GENCODE project aims to annotate with high accuracy based on computational, manual and experimental approaches. Manual annotations (HAVANA) are merged with automated annotations (Ensembl) with priority being given to the former. Non-coding transcripts are characterised as 'antisense', 'lincRNA', 'miRNA', 'mT_rRNA', 'mT_tRNA', 'rRNA', 'snoRNA' or 'snRNA' if well characterised or 'non_coding', 'processed transcript', 'retrotransposed' or 'misc_RNA' if poorly characterised. Annotations do not include MHC and immunoglobulin genes or single-exon genes. Annotations are categorised 1 to 5 (1 having the best support) or 'not analysed'. Confidence of annotation is based on mRNA and EST alignments from UCSC (see above) and Ensembl (see below). Support is also gained from the International Nucleotide Sequence Database Collaboration (GenBank, EMBL, DDBJ). Suspected erroneous transcripts, as identified by Ensembl, UCSC, HAVANA and RefSeq are flagged [488, 489].
NCBI RNA Reference Sequence Collection (RefSeq)
Comprehensive gene annotations are based on GenBank, NCBI computational analysis, user feedback and manual curation. Uncurated annotations are supported by WGS, inference by genome sequence analysis eg. Homology, computational predictions and protein homology support. Curated annotations have functional descriptive information added [490-492].
Vertebrate Genome Annotation (Vega) database
Long ncRNA (lncRNA) are defined as annotated genes with no ORF. These are subdivided: lncRNA non-coding: gene is known from published literature to not code for a protein; lncRNA 3prime_overlapping_ncrna: ditag and/or published experimental data supports the existence of short non-coding transcripts. lncRNA ambiguous ORF: transcripts are believed to be protein coding (take this into account in judging annotation evidence!!) with more than one possible ORF. lncRNA antisense: gene overlaps a coding exon on the opposite DNA strand or is known in the literature to regulate an antisense coding gene. lncRNA long interspersed ncRNA (lincRNA): Gene is intergenic to protein coding genes and is >200bp. lncRNA ncRNA_host: Gene has a separate transcript that hosts other ncRNAs such as microRNAs. lncRNA retained_intron: gene has alternatively spliced transcripts that contain intronic sequence relative to other coding transcripts. lncRNA sense_intronic: Gene has a long non-coding transcript in introns of a coding gene, which does not overlap coding exons. lncRNA sense_overlapping: On the same strand as a coding transcript that resides within an intron of the non-coding transcript. lncRNA processed_transcript: Doesn't contain an ORF but is miscellaneous to other lncRNA biotypes [493].
Ensembl Gene Predictions – Ensembl 68
Includes automatic and manual curation (by HAVANA) of confirmed and predicted non-coding and protein coding genes. LincRNA annotation is supported by presence of H3K4me3-K36me3 domains, limited ORF (<35%), lack of protein domain evidence (PFAM, InterPro) [494, 495].
lincRNA and Transcripts of Uncertain Coding Potential (TUCP)
Includes de novo annotations not covered by RefSeq, UCSC and GENCODE identified by RNA-Seq of 24 human cell subsets. Annotations have been filtered to remove transcripts with any evidence of protein domain or coding potential and non-lincRNA annotations. Transcripts of lower confidence are termed 'TUCPs' (Transcripts of uncertain coding potential) [496].
C/D and H/ACA Box SnoRNAs, ScaRNAs and MicroRNAs Identified with SnoRNABase and MiRBase
Covers annotations of pre-miRNA from miRBase [497] and snRNA and scaRNA from snoRNABase [498]. Presence of C/D and H/ACA are sequences indicating rRNA and snRNA modification activity.
Human expressed sequence tags (ESTs), spliced and unspliced
ESTs from Genbank ® [483]. Spliced ESTs have evidence of at least one intron of 32bp to 750kb and GT/AG ends.

A scoring system was developed to assess the priority of each transcript for future analysis Table 2.7.

Table 2.7: Scoring system for transcript prioritization

Score	*Criteria*	*Outcome*
0	At least one exonic protein-coding, miRNA, snRNA rRNA or tRNA annotation	Transcript dismissed
1	No annotation and no EST evidence or unspliced ESTs only	
2	No annotation and evidence of spliced EST	Transcript considered for future analysis
3	Annotated as non-coding by at least one consortia but with low confidence	
3	Annotated as non-coding by more than one consortia and with high confidence	Transcript retained for future analysis
5	Annotated as long ncRNA by all consortia with high confidence	
6	Annotated as long ncRNA by all consortia, with high confidence and has some background literature	Transcript prioritised for future analysis

2.5.6 Gene ontology

Gene ontology was performed using DAVID GO [499, 500] using Agilent Oligo IDs as identifiers. Unless otherwise specified, the background reference was all probes on the array.

2.5.7 Identification of disease associations with non-coding genes of interest

Information on published disease associations was taken from publically accessible databases: Genetic Association Studies of Complex Diseases and Disorders (GAD, [501], OMIM [502], ClinVar [503], the International Standards for Cytogenomic Arrays Consortium (ISCA, [504], the Catalogue Of Somatic Mutations In Cancer (COSMIC, [505], the Rat Genome Database (RGD, [506]), Human Gene Mutation Database Public Variants [507], Gene reviews [508], NHGRI catalog of published GWAS [509], and publications. These databases encompassed dbSNP entries.

2.5.8 Identification of transcription factors

Genes encoding transcription factors were identified by comparison with the most recently published peer-reviewed catalogue of human

transcription factors [510]. This catalogue is based on evidence for sequence specific DNA binding activity, excluding basal transcription factors that form the core transcriptional machinery. This evidence includes the presence of a known DNA binding domain and published experimental evidence for high confidence entries. Lower confidence transcription factors, for which no functional evidence exists for their activity, were included due to the presence of domains never observed in proteins without transcription factor activity.

2.5.9 RNA-Seq analysis

Sequencing data was received in FASTQ format. Adaptors and PCR duplicates were removed and paired-end unique reads above 20bp in length were aligned to reference genome hg19 with TopHat2 using default parameters (performed by John Ambrose). Bam and bigwig files were created with Samtools [511] and Bamtools [512].Cuffdiff2 [513] was used for differential expression analysis (John Ambrose, Alessandro Riccombeni). Read coverages were calculated using ngsplot and feature count (Bamtools, [512]. For statistical comparisons of RPKM the Kolmogorov-Smirnov Test was used [514]. This is non-parametric and distribution free. Data visualisation was performed in UCSC Genome Browser [482] and Microsoft Excel ®.

2.5.10 ChIP-Seq analysis

Sequencing data was analysed by Aditi Kanhere, Richard Jenner, Alessandro Riccombeni. Peak calling was performed with MACS (Parameters?) (Richard Jenner, [515]). T-bet stitched enhancers and super-enhancers were defined by Richard Jenner using the Rank Ordering

of Super-enhancers (ROSE) algorithm [320, 321]. Briefly, the ROSE algorithm scans 12kb windows across the genome, stitching together transcription factor peaks within these windows as a single enhancer. Enhancers are then ranked by RPKM of transcription factor reads. The cut-off for super-enhancer versus enhancer is taken at the point of inflection with normal enhancers having a linear relationship between rank and RPKM and super-enhancers having an exponential relationship. Enhancers within 2kb of a transcription start site are removed from analysis. H3K4me3 ChIP-Seq data were provided by Paul Lavender and P-TEFb and RNAPII ChIP-Seq experiments were performed by Richard Jenner and Arnulf Hertweck.

2.5.11 Publically available genomic data used in this work

Publically available genomic data used in this work is catalogued in Table 2.8. Quality and fidelity of data was assessed with FastQC [516].

Table 2.8: Publically available genomic data used in this work
Publically available RNA-Seq and ChIP-Seq data was used in some analyses presented here. These can be identified using the accession numbers provided here. WCE: Whole Cell Extract.

Species	***Cells***		***Data***		***Accession numbers and references***
Human	Treg	*Ex vivo* CD4+CD25+CD127-	ChIP-Seq	H3K4me3	GSM772944
				H3K4me1	GSM772973
				H3K27me3	GSM772946
				H3K9me3	GSM772943
				H3K36me3	GSM772945
				WCE	GSM772914
		Activated (CD3/28) CD4+CD25+		FOXP3	SRX060160 [517]
			RNA-Seq	PolyA+	SRX060156 [517]
		Ex vivo CD4+CD25+			SRX060154 [517]
	Tresp	*Ex vivo* CD4+CD25-	ChIP-Seq	H3K4me3	GSM772867
				H3K4me1	GSM772849
				H3K27me3	GSM772868
				H3K9me3	GSM772850
				H3K36me3	GSM772866
				WCE	GSM772915
		Activated (CD3/28) CD4+CD25-		FOXP3	SRX060163 [517]
			RNA-Seq	PolyA+	SRX060159 [517]
		Ex vivo CD4+CD25-			SRX060158 [517]

3 Identification and characterization of ncRNAs upregulated in Treg

3.1 Introduction

3.1.1 T cell lineage plasticity

CD4 T cell lineages play specialised roles within the immune system with Th1 versus Th2 phenotypes, which are tailored towards type I and type II infections, respectively, epitomising this. While there are distinct characteristics to each lineage (Section 1.1), for example the production of lineage-specific cytokines, there is inherent flexibility. This plasticity is shown through the co-expression of lineage-specifying transcription factors, which were once thought of as master regulators (Section 1.2.1), and the bivalent presence of permissive and repressive histone modifications at lineage-specific genes, which allows repressed genes to be maintained in a poised-to-be-active state (Section 1.2.3). This plasticity has functional relevance in facilitating the immune response to a changing environment but there are deleterious consequences when lineage choice becomes dysregulated. In the systemic autoimmune disease SLE there is some evidence of Treg abnormalities, particularly when naïve and memory subsets, defined by differential expression of CD45RA, are observed. These defects could relate to disease aetiology or could be the result of the disease environment. Understanding the perturbation of Treg phenotype in SLE at a transcriptional level will inform models of Treg and T cell plasticity and may give insight into disease pathogenesis.

3.1.2 ncRNAs show lineage-specific expression

Multiple classes of ncRNAs have been reported to exhibit cell type specific expression (Section 1.3.2). Several mechanisms have been suggested for

how these ncRNAs may influence cell identity (Section 1.3.3) including the possibility that they may influence the expression of other lineage-specific genes at the level of transcription through the recruitment of chromatin modifying complexes. The functional importance of several ncRNAs has been demonstrated. However, to date no non-coding RNAs have been characterised in Treg.

3.1.3 Aims

1. Define a specific set of lincRNAs upregulated in Treg compared to Tresp.
2. Characterise these Treg-specific lincRNAs.
3. Examine differential expression of mRNAs and incRNAs between healthy individuals and SLE patients.
4. Identify differences in gene expression and histone modification between naïve and memory subsets of Treg and between SLE patients and healthy controls.

3.2 Results

3.2.1 Identification of differentially expressed transcripts between Treg and Tresp

I began investigating the transcriptional differences between Treg and Tresp by examining gene expression differences between these populations, focusing on naïve and memory subsets as these have been demonstrated to exhibit differing phenotypes with Treg, which are not fully understood. To achieve this, naïve and memory Treg and Tresp were isolated as described in Sections 2.2.1 to 2.2.6 with donors ranging in age from 23 to 54 years (median: 30; standard deviation: 12) and being female. RNA was prepared according to Sections 2.3.1 to 2.3.3 and microarrays performed as per Section 2.3.5.. Array data were extracted and normalised as per Section 2.5.1. To gain an overview of the data I performed hierarchical clustering and visualised correlations in gene expression using Treeview, as described in Section 2.5.2. This showed trends in gene expression seen in previously published research (Figure 3.1), including expression of the genes associated with Treg function such as those encoding for the lineage-specifying transcription factor, FOXP3 [37-39], the IL-2 cell surface receptor, IL2RA (CD25) [33], the CTLA4 cell surface receptor, which is constitutively expressed on Treg and mediates suppressive function [518] and the transcription factor, IKZF2 (Helios) which has been identified as being upregulated in Treg but whose function is not yet well understood [93]. The gene for the cell adhesion molecule PECAM1 (CD31), which is known to be expressed upon recent thymic emigrants [519], was upregulated in naïve cells compared to memory for both Treg and Tresp, as expected. Published data on differential

expression of AHR and RORC between naïve and memory Treg [96] were also recapitulated. These reflections of previously published data gave confidence that this microarray data can be used to detect changes in gene expression relevant to Treg biology. To identify genes with statistically different expression between Treg and Tresp and naïve and memory subsets I used rank product analysis as described in Section 2.5.4. Figure 3.2 illustrates the numbers of significantly up- and downregulated genes unique to each population. The identity of these transcripts is detailed in Section 7.1 of the appendix. Figure 3.2.A shows that 527 transcripts in total were differentially regulated between all Treg and all Tresp with 320 (61%) being upregulated in Treg. Fewer differences in gene expression were identified between naive and memory subsets of Treg and Tresp (Figure 3.2.B). For naïve Treg no transcripts were identified as uniquely downregulated, for naïve Tresp only 2 were identified as upregulated and for Tresp memory, 6 (2%) were identified as downregulated. More transcripts were uniquely upregulated in naïve Treg than to naïve Tresp (70 versus 2, respectively). This could reflect a more specialised role for naïve Treg, which exit the thymus as a distinct lineage, than naïve Tresp, which are undifferentiated and may be relatively transcriptionally inactive.

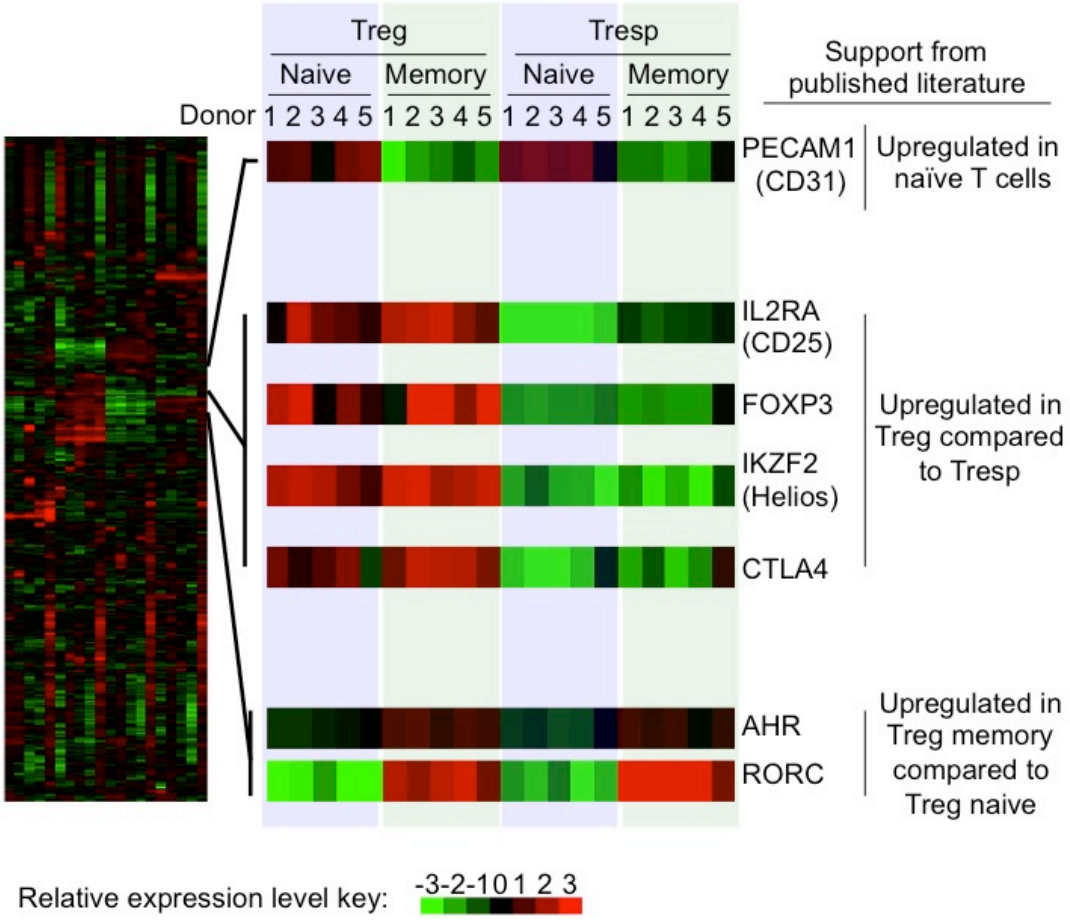


Figure 3.1: Gene expression patterns reflect published literature.
Correlations in gene expression between naïve and memory subsets of Treg and Tresp from healthy individuals were examined. These patterns are visualised here as a heatmap with reference to genes that demonstrated previously published gene expression trends.

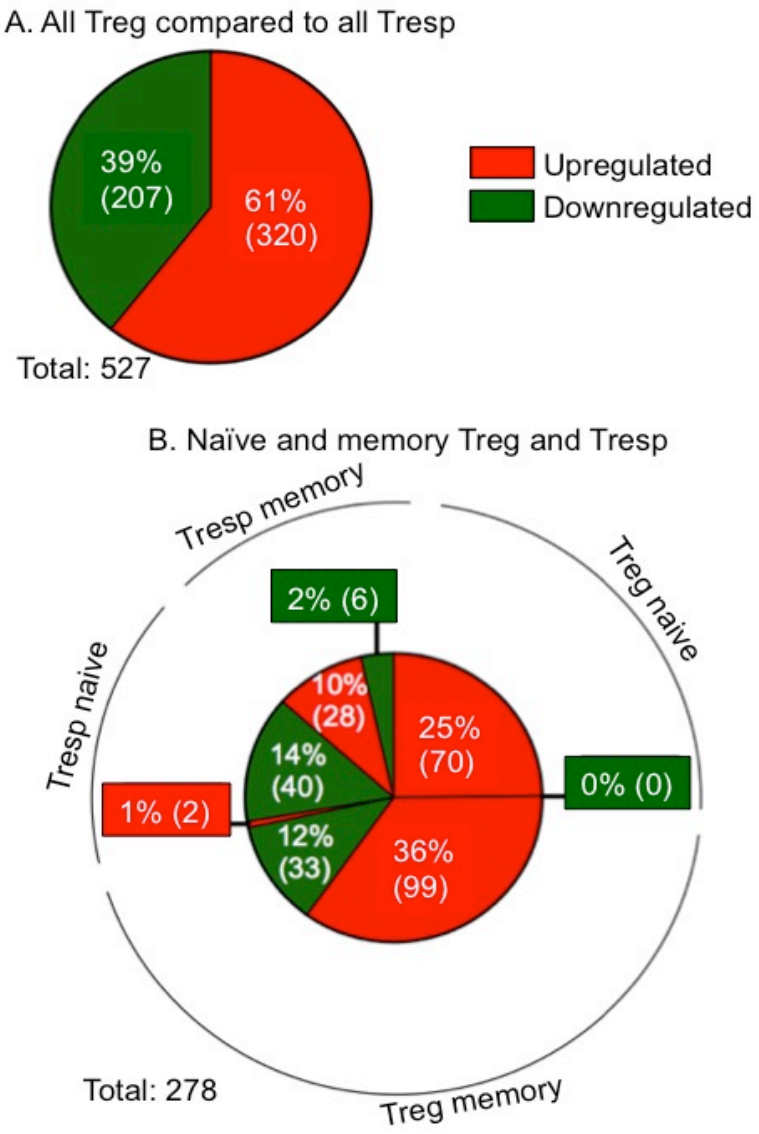


Figure 3.2: Numbers of differentially expressed transcripts from CD4 T cell populations from healthy individuals

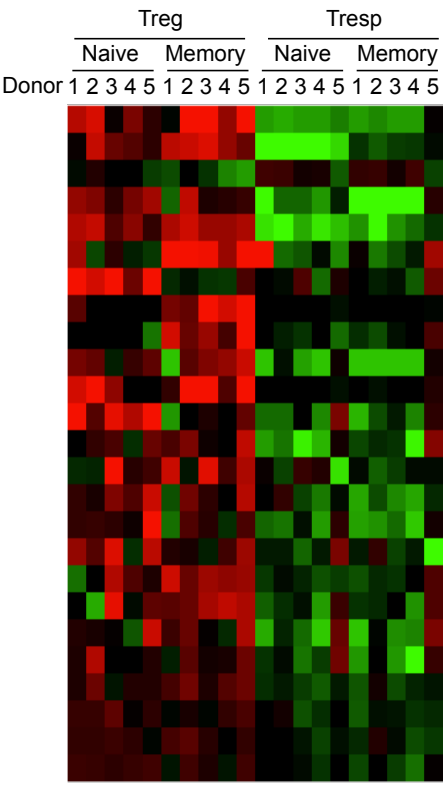
Transcripts with statistically significant differential expression were identified by Ranksum. A: Comparison of all Treg compared to all Tresp. B: Transcripts uniquely differentially regulated in each naïve and memory subset. The percentage of the total number of differentially expressed transcripts is given while bracketed numbers indicate the absolute number of transcripts.

3.2.2 Identification and validation of non-coding RNAs upregulated in Treg

The process of distinguishing non-coding from coding transcripts relied upon publically available gene annotation evidence, as described in Section 2.5.5. Of the 320 transcripts upregulated in Treg, 22 (6.9%) were categorised as non-coding. This proportion does not represent relative biological ratios of protein coding to non-coding transcripts but instead reflects array design bias and the stringency of non-coding characterisation, as described in Section 2.5.5. The identities of these transcripts are given in Table 7.25 of the appendix and their relative expression levels are illustrated in Figure 3.3.

I next sought to validate the differential expression of these ncRNAs in an independent set of donors. Of these non-coding transcripts upregulated in Treg compared to Tresp, differential expression of 7 (29%) was validated by qPCR in three additional donors (Figure 3.4): MIR146A (fold difference: 21.8; p=0.177), CRNDE (fold difference: 7.2; p=0.0178), LINC00312 (fold difference: 20.8; p=0.0183), PTTG3P (fold difference: 2.12; p=0.006), LOC286442 (fold difference: 7.0; p=0.011), ENST00000444919 (fold difference: 1.4; p=0.0508), ENST00000415387 (fold difference: 43.8; p=0.0137). The array probe and qPCR primers detect the pri-miRNA form of miR146a, which has been previously identified as upregulated in Treg before [520] and has functional relevance for Treg biology [521]. Published work on miR-146a in Treg has been limited to differential expression of the mature form of the miRNA, so detection of the pre-processed form shown here indicates that miR-146a expression is regulated prior to activity of

miRNA processing machinery such as Dicer and Drosha. In addition, the ability to identify a Treg relevant miRNA supported the approach used here for finding additional ncRNAs.



	Fold difference (Treg/Tresp)	PFP
FOXP3	11.34	0.00
IL2RA (CD25)	13.80	0.00
IL7R (CD127)	0.57	1.02
LOC286442	13.89	0.00
MIR146A	13.40	0.00
LINC00312	9.41	0.11
LOC283663	8.73	0.26
PMCHL1	5.59	0.05
lincRNA:chr3:8615882-8616354_F	5.30	0.11
ENST00000432978	5.14	0.00
ENST00000415387	4.91	0.00
ENST00000411475	3.91	0.01
lincRNA:chr13:50614546-50615364_R	3.57	0.00
CRNDE	3.42	0.04
LOC100131234	3.40	0.00
SLC9A7P1	3.21	0.00
lincRNA:chr2:64455559-64478475_F	3.08	0.03
PTTG3P	2.94	0.01
lincRNA:chr7:7928975-7996750_R	2.87	0.03
ZNRF2P1	2.80	0.03
ENST00000444919	2.45	0.04
LOC100132741	2.42	0.00
ENST00000502284	1.82	0.04
lincRNA:chr3:106619835-106630535_R	1.80	0.04
A_33_P3332616	1.80	0.04

Figure 3.3: Heatmap illustrating statistically significant upregulation of ncRNAs in Treg compared to Tresp

Twenty-two Treg upregulated transcripts were identified as ncRNAs. They are shown here in order of fold difference between Treg and Tresp. The predicted false positive rate (PFP, analogous to a false discovery rate) is also given to demonstrate the statistical significance of their differential expression. The genes FOXP3 and CD25, which are known to be upregulated in Treg, are given as a comparison, as is CD127, which is downregulated in Treg,

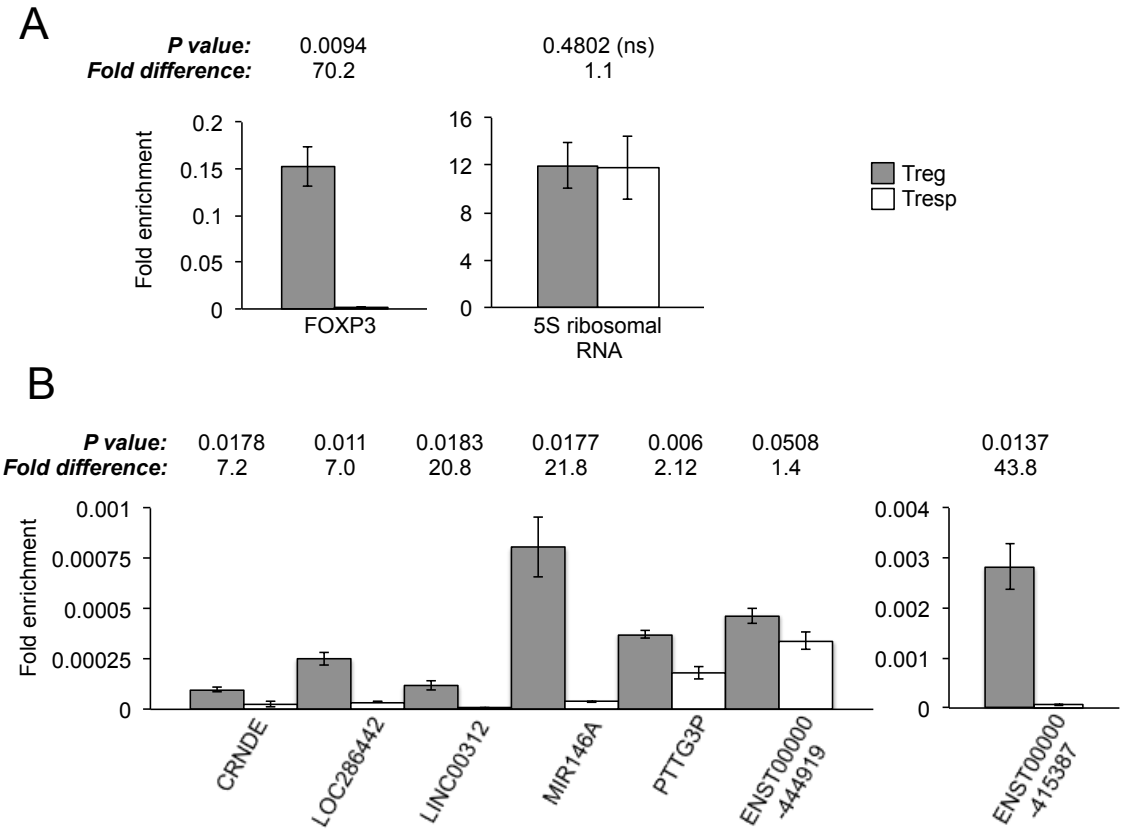


Figure 3.4: QPCR validation of ncRNA expression

Quantitative PCR in additional donors to those used to obtain microarray data. A: FOXP3 (left), is known to be upregulated in Treg compared to Tresp while the housekeeping gene 5S ribosomal RNA (right) should be similarly expressed between the two populations. B: Treg upregulated ncRNAs identified by array that were validated by qPCR. N=3; Error bars represent SEM, one-tailed t-test.

3.2.3 Histone modifications at Treg ncRNA genes

To further validate the expression data, differential histone modifications between Treg and Tresp were examined using publically available ChIP-Seq data detailed in Section 2.5.11. The presence of H3K4me3 (a mark of transcriptional initiation) at the TSS and H3K36me3 (transcriptional elongation) across the gene body would support the upregulation of these ncRNAs at the transcriptional level. The presence of H3K4me3 at an ncRNA TSS supports the assumption that the ncRNA derives from an independent gene rather than being a product of RNAPII read through from a neighbouring gene.

The relative enrichment of these histone modifications between Treg and Tresp supports the regulation of these genes at the level of transcriptional, rather than through post-transcriptional mechanisms, such as targeted RNA degradation, and can also be used to determine if this regulation is independent of neighbouring genes or is a by-product and possibly 'transcriptional noise'.

Figure 3.5 shows the enrichment of each of these histone modifications at the ncRNA genes in primary human Treg and Tresp. As expected, the Treg upregulated gene *FOXP3* shows enrichment of H3K4me3 and H3K36me3 in Treg while *IL2*, which is not expressed in Treg, shows greater enrichment of H3K4me3 in Tresp and the housekeeping gene *ACTB* shows no cell type specific enrichment of either modification. Of the Treg upregulated ncRNAs, *LOC286442* and *MIR146A* both show greater enrichment of H3K4me3 and H3K36me3 in Treg. These were also the most overexpressed Treg ncRNAs in the microarray data (Figure 3.3). All

other Treg ncRNAs do not show obvious differential enrichment between Treg and Tresp but *CRNDE* is marked with H3K4me3 in both cell types, as is *ENST00000444919*. Several ncRNA genes do not appear, by histone modifications, to be transcribed independently of their neighbouring genomic genes. For example, *LINC00312* has greater H3K36me3 enrichment in Treg but this may be due to regulation of the upstream gene *LMCD1*, which is supported by the lack of H3K4me3 at *LINC00312*. Likewise, the slight increase in H3K36me3 over *ENST00000415387* in Treg over Tresp and the lack of H3K4me3 may reflect expression of the upstream gene *IKZF2* (Helios), which has a known Treg specific expression pattern. In addition, the ENST00000415387 gene overlaps with *IKZF2* and so transcription from this region may be due to 'read-through' of *IKZF2* rather than active regulation of *ENST00000415387*.

The presence of monomethylated of H3K4 (H3K4me1) has been associated with poised and active enhancer regions [289, 344, 356] where it is enriched relative to H3K4me3. While both marks are often found at the TSS of protein coding genes and reflect how these genes are regulated, a variation in pattern is suggestive of regions with enhancer activity. In addition, enhancer regions have been shown to produce ncRNA, which may play a role in functional effects on the transcriptional regulation of other genes. Therefore examining the presence of H3K4me1 at the genomic loci of Treg ncRNA can inform their potential role in transcriptional regulation as enhancer RNAs. As

Figure 3.5 illustrates, the *FOXP3* gene shows prominent enrichment of H3K4me3 at the TSS while H3K4me1 is minimal in comparison. *MIR146A*

shows a similar pattern as does *CRNDE*, which suggests these are regulated in a similar way to protein coding genes as is expected for lincRNAs and pri-miRNAs. *PTTG3P* however shows little enrichment of neither H3K4me3 nor H3K4me1. *ENST00000444919* has greater enrichment of H3K4me3 compared to H3K4me1 at its TSS and in addition has shows some H3K4me1 at its gene body, as does *ENST00000415387* however this could relate to the presence of their overlapping neighbouring genes, *ITGA6* and *IKZF2*, respectively. H3K4me1 is also seen along *LOC286442*, which is not overlapped by an annotated gene. *LINC00312* is most 'enhancer-like' showing poor enrichment of H3K4me3 but enrichment of H3K4me1 nearby. In addition, *LINC00312* is monogenic, which is a feature of enhancer RNAs.

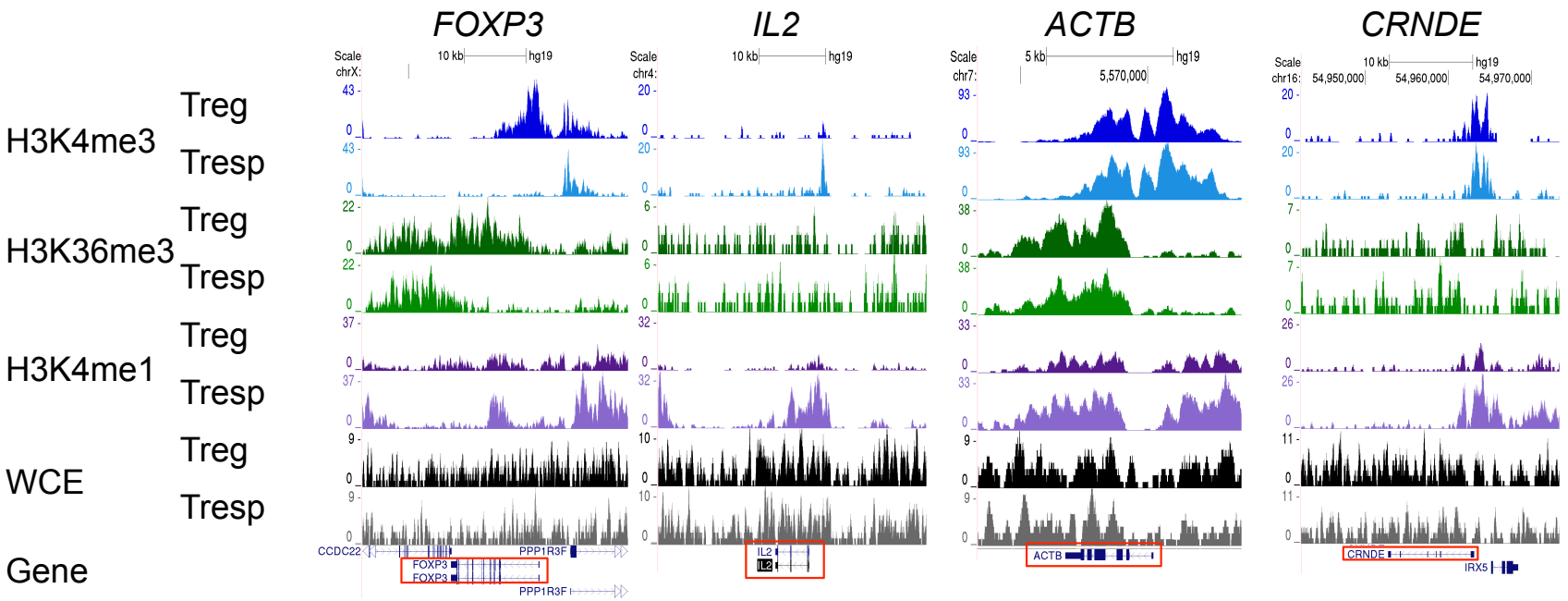


Figure 3.5: Histone modifications at ncRNA genes upregulated in Treg

(Continued to page 134) The relative enrichment of histone modifications associated with transcriptional initiation (H3K4me3) and elongation (H3K36me3) and enhancer regions (H3K4me1) at Treg ncRNA genes in Treg and Tresp. WCE represents the background total DNA. *FOXP3* and *IL2* genes are given as examples of genes known to be upregulated in Treg and Tresp, respectively, while the housekeeping *ACTB* is present at similar levels in both cell types. H3K4me3: Trimethylation of lysine 4 on histone 3; H3K36me3: Trimethylation of lysine 36 on histone 3; H3K4me1: Monomethylation of lysine 4 on histone 3; WCE: Whole cell extract.

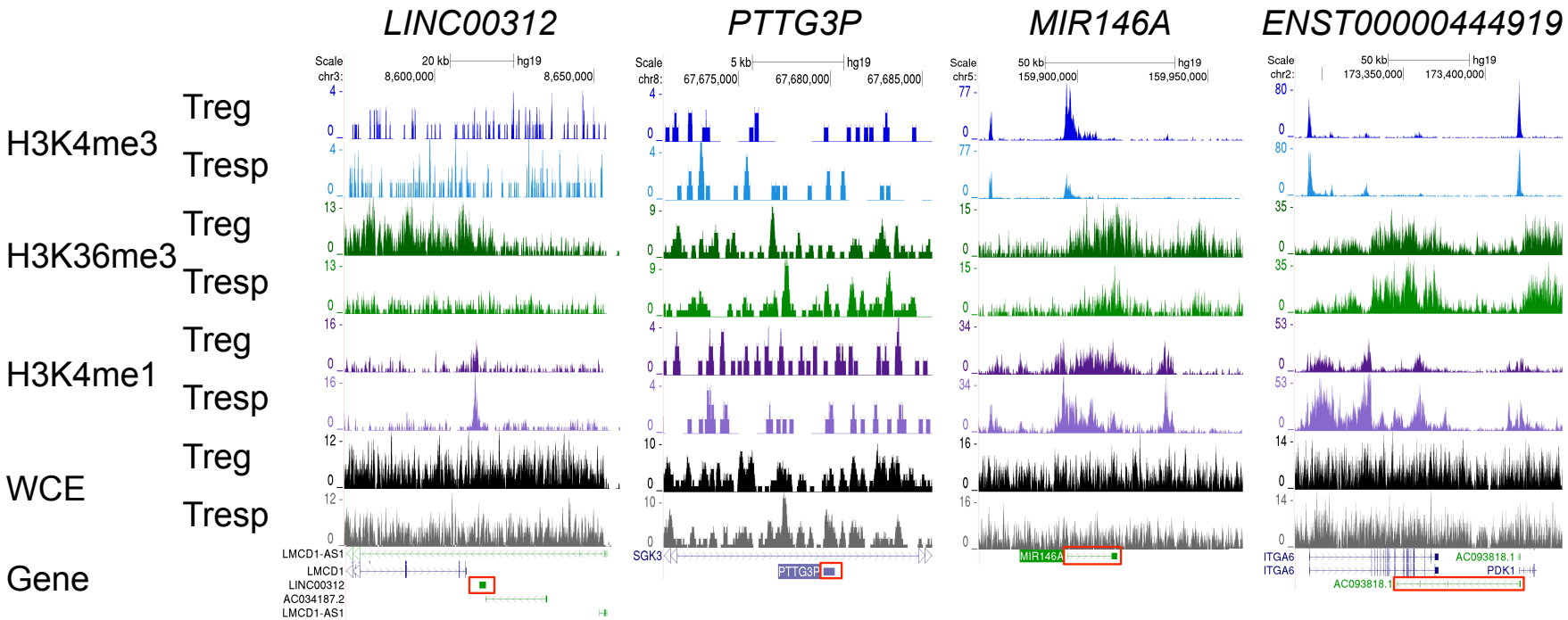


Figure 3.5: Histone modifications at ncRNA genes upregulated in Treg

(Page 132-134) The relative enrichment of histone modifications associated with transcriptional initiation (H3K4me3) and elongation (H3K36me3) and enhancer regions (H3K4me1) at Treg ncRNA genes in Treg and Tresp. WCE represents the background total DNA. *FOXP3* and *IL2* genes are given as examples of genes known to be upregulated in Treg and Tresp, respectively, while the housekeeping *ACTB* is present at similar levels in both cell types. H3K4me3: Trimethylation of lysine 4 on histone 3; H3K36me3: Trimethylation of lysine 36 on histone 3; H3K4me1: Monomethylation of lysine 4 on histone 3; WCE: Whole cell extract.

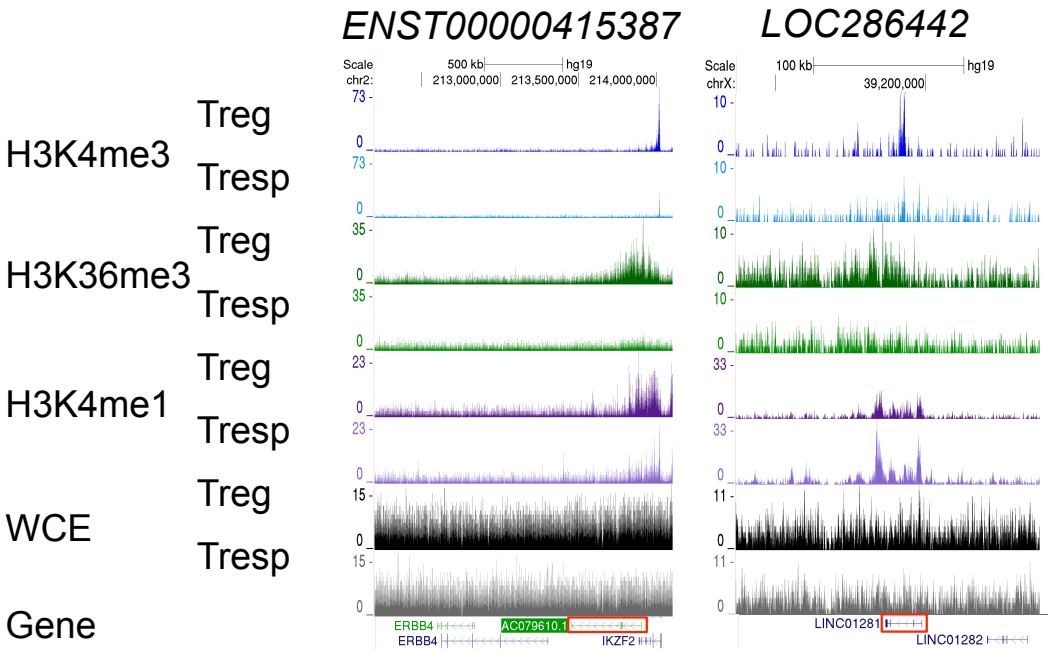


Figure 3.5: Histone modifications at ncRNA genes upregulated in Treg

(Continued from page 132) The relative enrichment of histone modifications associated with transcriptional initiation (H3K4me3) and elongation (H3K36me3) and enhancer regions (H3K4me1) at Treg ncRNA genes in Treg and Tresp. WCE represents the background total DNA. *FOXP3* and *IL2* genes are given as examples of genes known to be upregulated in Treg and Tresp, respectively, while the housekeeping *ACTB* is present at similar levels in both cell types. H3K4me3: Trimethylation of lysine 4 on histone 3; H3K36me3: Trimethylation of lysine 36 on histone 3; H3K4me1: Monomethylation of lysine 4 on histone 3; WCE: Whole cell extract.

3.2.4 FOXP3 binding at Treg ncRNA genes

Lineage-specifying transcription factors orchestrate the transcriptional program of each T cell lineage. The corresponding transcription factor for Treg, FOXP3, has been associated with both up and down regulated genes in Treg. For example, it is seen at the IL-2 gene, which encodes a major T cell growth factor and is therefore not expressed by anti-inflammatory Treg, and at the FOXP3 gene itself, implying it enforces its own expression through positive feedback. FOXP3 has therefore been suggested to act as either a repressor or activator, depending on context. Using publically available ChIP-Seq data from human Treg (detailed in Section 2.5.11), I investigated whether the genes encoding the Treg upregulated ncRNAs were bound by FOXP3. If so, this would support the Treg specific biological relevance of these ncRNAs. As Figure 3.6 shows, FOXP3 bound its own gene at the promoter and first intron, as previously shown in mouse [258] and the miR-146a gene which has a known role in Treg biology is also bound, which has not be published. Of the Treg ncRNAs, *CRNDE* and *ENST00000444919* exhibit FOXP3 binding at their promoters, which they share with other genes and may indicate the ncRNA genes are not specifically targeted by this FOXP3 recruitment. LINC00312 does not appear to be bound by FOXP3 although the upstream gene *LMCD1* is, which supports the possibility that *LINC00312* is transcribed as a byproduct of *LMCD1* activation. *ENST00000415387*, also has FOXP3 binding in its 3' region, which overlaps IKZF2, but this binding also extends further along the gene in an intronic region. The FOXP3 binding seen near *PTTG3P* this could be targeting the intron of the gene antisense to *PTTG3P*. *LOC286442,* which is relatively distant to

other annotated genes, has FOXP3 binding at both its promoter and in an intronic region.

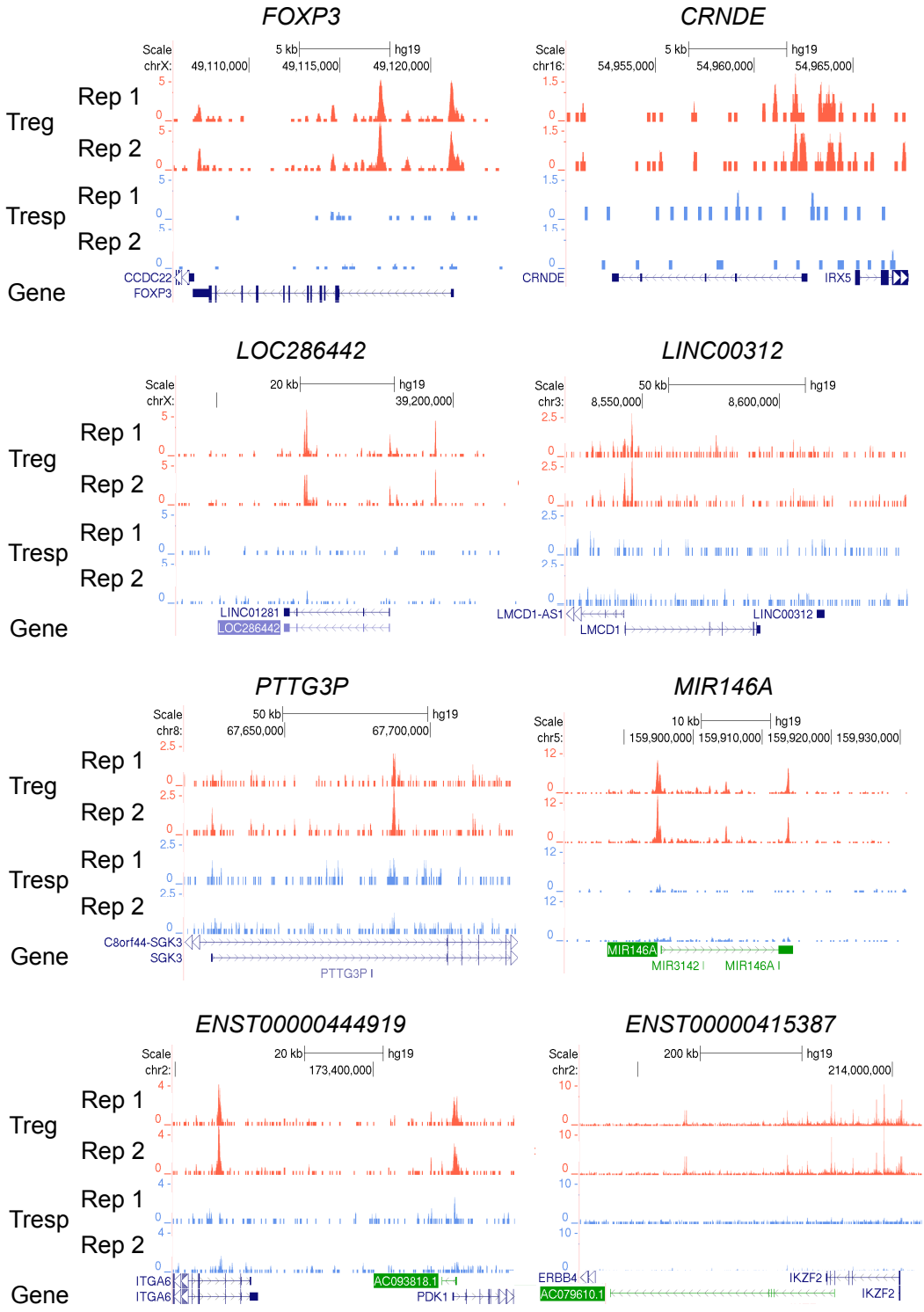


Figure 3.6: FOXP3 binding at Treg ncRNA genes

FOXP3 occupancy at Treg ncRNA genes. Rep: replicate. ChIP-Seq was obtained from a publically available resource detailed in Section 2.5.11.

3.2.5 Genomic structure of Treg ncRNA genes

Being defined by a lack of protein coding capacity in their primary sequence, the identity and role of ncRNAs is underpinned by their genomic evolution. Understanding the biological function of ncRNAs can therefore be informed by knowledge of their genomic structure. Figure 3.7 illustrates the genomic location of each validated Treg ncRNA and displays intron-exon structure, alternative isoforms and repetitive elements.

CRNDE is notable for its many (10) alternative isoforms compared to the other ncRNAs that have at most 2 (LOC286442) but commonly only one isoform. In terms on intron-exon structure, LINC00312 and PTTG3P consist of one exon without any intronic sequence while LOC286442, MIR146A, ENST00000444919 and ENST00000415387 contain large intronic regions.

Repetitive elements are more associated with non-coding genes than protein coding genes [404], possibly because their mechanisms of function are more tolerant to transposon insertion i.e. there are no open reading frames or stop or start codons to disrupt. It has also been suggested that repeats may contribute to ncRNA function [404] as they would allow base pairing for self-annealing, leading to the formation of potentially functional secondary structures, and annealing with other ncRNAs. These functional repeat-sequence domains are seen in Kcnq1ot1 [522] and Xist [523]. The repeats seen in CRNDE correspond with exonic sequences (Figure 3.7), which would be consistent with them

having a role in the transcribed product. LINC00312 and PTTG3P are notable by their lack of repeats. This is consistent with PTTG3P, a pseudogene, diverging from a protein-coding gene and might suggest LINC00312 has a similar evolutionary origin. The other ncRNA genes (LOC286442, MIR146A, ENST00000444919 and ENST00000415387) contain several repetitive elements of various classes. The presence of repeats in lincRNAs has been associated with expression in other contexts, for example human endogenous retrovirus subfamily H (HERVH) containing lincRNAs appear to transcribed in a stem cell specific manner [407].

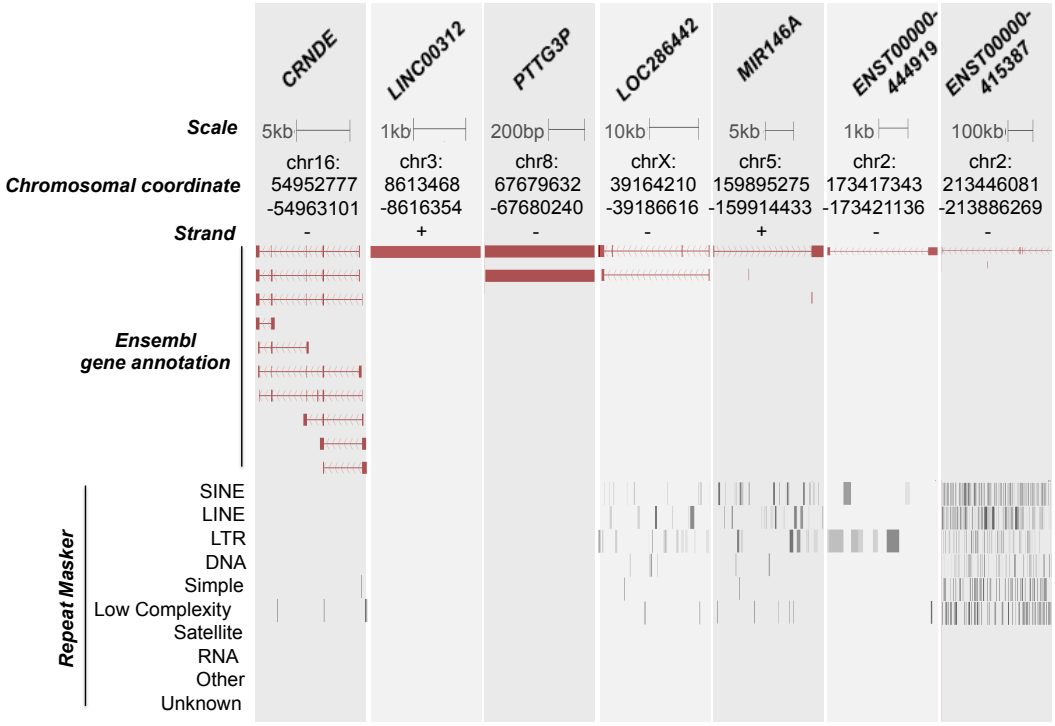


Figure 3.7: Genomic identities of ncRNA genes upregulated in Treg

The genomic structure, location and repetitive character of each Treg ncRNA gene. Ensembl annotations are shown here, as these were most comprehensive. Alternative isoforms are shown as well as any intronic, same strand transcripts annotated as other genes. For simplicity, antisense annotations have been removed from the view. SINE: short interspersed nuclear elements; LINES: long interspersed nuclear elements; LTR: long terminal repeats.

3.2.6 Correlations between Treg ncRNA expression and genomic neighbours

Several mechanisms have been suggested for how ncRNAs may influence cell phenotype. This includes at the level of transcriptional regulation of other genes with known biological functions. This influence of ncRNAs in transcriptional regulation has been seen in *cis* at genes encoded in close proximity to ncRNA genes (Section 1.3.3.1, [403]). Therefore Treg ncRNAs encoded near genes with known biological functions in Treg, T cells or with relevant biological processes are of particular interest. I examined the nearest gene annotations upstream and downstream of each validated ncRNA genes (Figure 3.8). Their orientation relative to the ncRNA gene was also noted as this impacts on the mechanism of cis-regulatory ncRNAs (Section 1.3.3.1). I also correlated the expression of each ncRNA and its neighbouring genes (Figure 3.9).

CRNDE is encoded 2kb from the TSS of *IRX5* in a divergent antisense orientation (Figure 3.8). This orientation suggests that the two genes may share promoters or other regulatory regions and published research on promoter characteristics in this area, which includes a CpG island, supports this [524]. Concordant expression between the two genes has been previously noted [525, 526] although expression of *CRNDE* without *IRX5* has also been seen [526]. *IRX5* encodes an Iroquois homeobox transcription factor with roles in development of several tissues [527, 528] however there is no literature relating to the role of IRX5 in T cells. Examining *CRNDE* and *IRX5* across naïve and memory Treg and Tresp

shows a relatively poor correlation in expression between the two genes (Pearson correlation coefficient: 0.49; Figure 3.9). The nearest gene encoded downstream of *CRNDE* is another ncRNA (RP11-1136G4.1). The product of this gene is poorly characterised and so correlation in expression would not inform the function of CRNDE. Co-expression data is also limited as no probes to this gene were incorporated into the array design.

LINC00312 is overlapped by another ncRNA gene, *AC034187.2*, which is poorly characterised and is not included in the array data. Also nearby are *LMCD1* and another ncRNA that is antisense to *LMCD1*, *LMCD1-AS1*. LINC00312 shows a relatively strong correlation in expression with *LMCD1* (Pearson correlation coefficient of 0.69, Figure 3.9) and both are preferentially upregulated in memory Treg above other subsets (Figure 3.9). No published data could be found on the LMCD1 (LIM and Cysteine rich domains 1) protein in T cells but in cardiac myocytes LMCD1 expression correlates with increased activity of calcineurin and NFAT [529], two signaling proteins that are relevant to T cell activation downstream of TCR signalling.

PTTG3P is antisense to *SGK3* (serum/glucocorticoid regulated kinase family, member 3), which is associated with non-T cell specific cell metabolism, proliferation and survival [530]. *VCPIP1* (valosin containing protein complex interacting protein 1) is located 100kb downstream of *PTTG3P* on the same strand and although abnormal CpG island methylation has apparently been seen in breast cancer [531] its biological

function is not well understood. *MCMDC2* (minichromosome maintenance domain containing 2) is encoded ~100kb upstream on the same strand as *PTTG3P* and again its function is poorly characterised and no probes for this gene are found on the array. Correlations in expression between *PTTG3P* and both *SGK3* and *VCPIP1* are weak (-0.33 and 0.34, respectively, Figure 3.9).

The gene encoding LOC286442 is located in a large region (hundreds of kbs) with no protein coding genes. As Figure 3.8 shows, there are however non-coding genes relatively nearby but the role of these is unknown and probes for these are not present on the arrays.

MIR146A shows a relatively strong co-expression (Pearson correlation coefficient of 0.78, Figure 3.9) with PTTG1 (pituitary tumor-transforming 1) that is oncogenic in several cellular contexts, including in T cells [532]. *PTTG1* is encoded upstream of *MIR146A* on the same strand. PTTG1 might also be implicated in T cells by its association with the Treg inducing cytokine TGF-β [533,], which has been observed in the context of breast cancer. ATP10B (ATPase, class V, type 10B) is encoded downstream of miR-146a but expression between the two is not strongly correlated and the role of this protein is not well understood.

ENST00000415387 is encoded near several other ncRNAs (Figure 3.8) but probes to these are not available on the array. Data on the protein coding genes *ERBB4* and *IKZF2* are available and of these the latter is strongly correlated with *ENST00000415387* (correlation coefficient: 0.73).

IKZF2 encodes the Treg specific transcription factor, Helios [93]. Helios has been suggested to distinguish nTregs from iTregs [89], but this assertion is controversial and the contribution of Helios to Treg function is still not understood. The relative orientation of the two genes, with the 5' end of *ENST00000415387* overlapping the 3' end of *IKZF2*, (Figure 3.8), is consistent with transcription read through from *IKZF2*.

Finally, *ENST00000444919* and *PDK1* are in an divergent orientation with their theoretical promoter regions occupying the same region (Figure 3.8), which would indicate their expression may be positively correlated. Unlike *CRNDE* and *IRX5*, which are in a similar orientation (see above), *ENST00000444919* and *PDK1* (pyruvate dehydrogenase kinase, isozyme 1), the mitochondrial enzyme that catalyses oxidative decarboxylation of pyruvate to acetyl coenzyme-A for entry into the Krebs cycle, are not strongly correlated (Pearson correlation coefficient: 0.49, Figure 3.9). *ITGA6*, an integrin that is expressed on T cells [534], is also on the opposite strand to *ENST00000444919* but in a convergent orientation (Figure 3.8) and also does not show a correlated expression pattern (Figure 3.9).

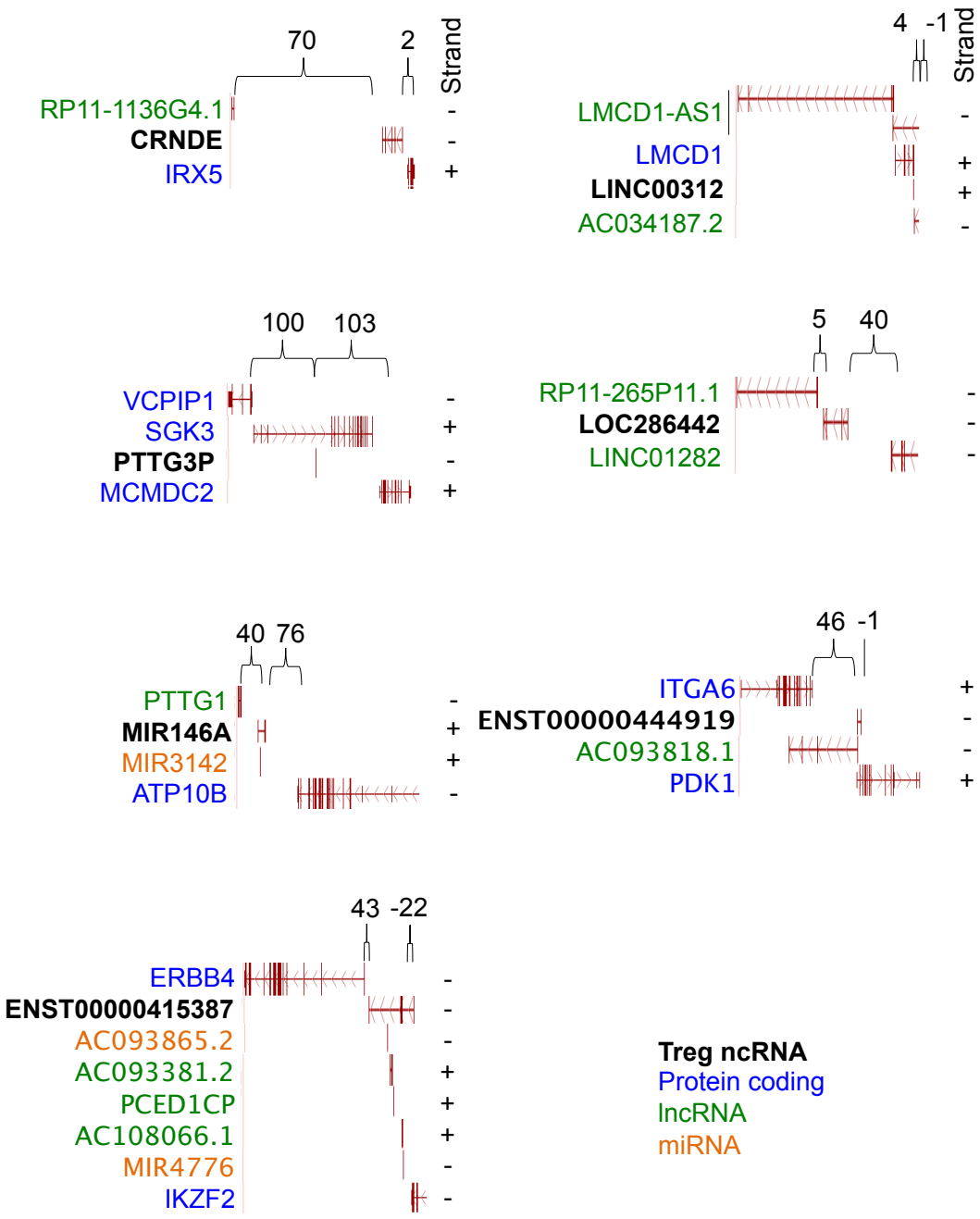


Figure 3.8: Genomic relationships of Treg ncRNA to other genes
The orientations of Treg ncRNA genes relative to neighbouring genes. Numbers indicate the distance (kb) between genes. A negative number indicates an overlap of genes. Black: ncRNAs; Blue: protein coding genes; Yellow: miRNAs; Green: other ncRNAs.

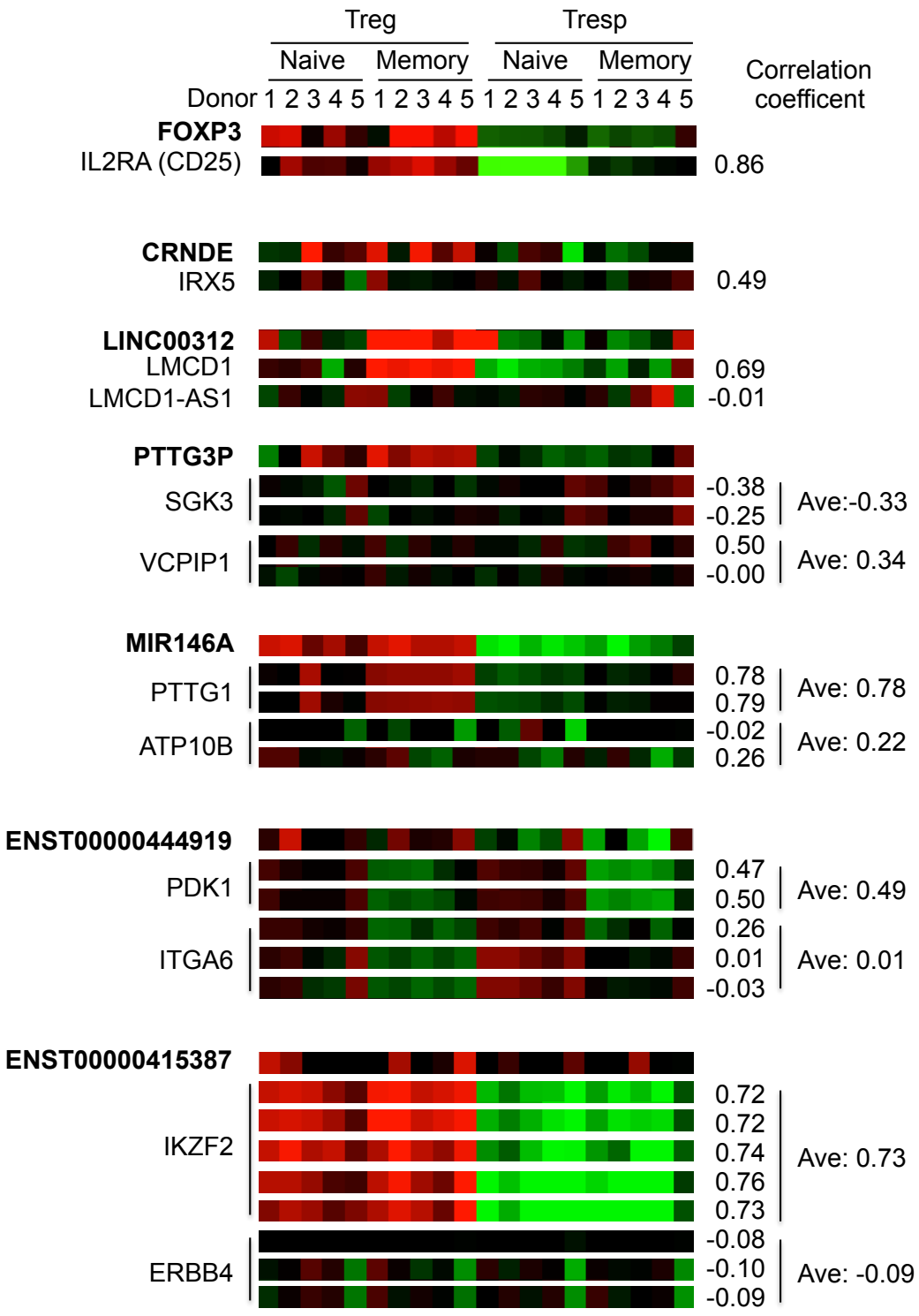


Figure 3.9: Correlation in expression of Treg ncRNA and genomic neighbours
Correlations in expression patterns between Treg ncRNAs and their neighbouring genes identified in Figure 3.8. Only neighbouring genes for which there were array probes are shown. Genes detected by several microarray probes have data for each probe presented separately. Co-expression was assessed by Pearson correlation coefficient with a cut-off for strong positive correlation taken at 0.7 and for a strong negative correlation at -0.7. The correlation between *FOXP3* and *IL2RA* (CD25) is shown as positive control...

3.2.7 Disease associations of ncRNA loci

Association of genomic variants in ncRNA loci with disease may reveal roles for these ncRNA in the biological processes underpinning these diseases. Although sequence variation in ncRNA does not have the same functional consequences as its does in protein coding genes, many trait associated variants lie outside of protein-coding regions [509], supporting the biological relevance of non-coding regions. At the DNA level, these non-coding regions may represent regulatory elements, such as enhancers and promoters, the mutation of which could influence disease through disrupted binding of regulatory proteins with deleterious effects on gene expression. At the RNA level, variation in non-coding regions may influence the activity of enhancers, which may mediate their long-range effects through RNA. In lincRNAs, these sequence variations could be deleterious through effects on the formation of RNA secondary structures and on the binding of proteins whose activity they influence. I used publically available data on the association of several classes of sequence variation with disease (Section 2.5.7) to inform the biological activity of these Treg ncRNAs. Accession numbers for relevant variants found and details on their disease associations are given in Table 3.1.

As Table 3.1 shows, every Treg ncRNA locus included an annotated disease-associated variation. The majority of these were copy number variations (CNVs, denoted by the ‘nsv-’ prefix to the accession number in Table 3.1) and several quantitative trait loci (QTLs) were also identified (‘RGD-’ prefix). All copy number variations and QTLs extended for several Mbp and included multiple gene annotations. As these disease

associations are not gene specific it is not likely that these ncRNA are involved in the processes underlying these diseases and therefore these categories of variants are not particularly informative of ncRNA function. In addition, these CNVs and QTLs are associated with non-specific descriptions of traits such as 'Developmental delay and/or other significant developmental or morphological phenotypes'. Single nucleotide polymorphisms (SNPs, denoted by the 'rs-' prefix in Table 3.1) however only span a single nucleotide, as the phrase implies, and so are localised to the ncRNA gene of interest. SNPS were seen in two Treg ncRNA genes: *MIR146A* and *ENST00000415387.* The *MIR146A* genomic sequence contains six SNPs in total, which are associated with several cancers and the autoimmune disease systemic lupus erythematosus (Table 3.1). Meanwhile, *ENST00000415387* contains a SNP associated with age-related cognitive decline. No disease causing (as opposed to disease associated) mutations were found at any of these non-coding loci.

Of the diseases associated with Treg ncRNA loci, listed in Table 3.1, those with an immunological basis and potential Treg involvement include the autoimmune rheumatoid diseases rheumatoid arthritis (RA), systemic lupus erythematosus (SLE) and cancer. The relationship between these ncRNA loci and these diseases may inform understanding of the activity of these ncRNA within Treg. Additional detail on these variants, including the statistical significance of disease variation and relevant literature are given in Table 3.2. The QTL covering LINC00312 is significantly associated with is one study of familial RA in North American probands

(p=0.0139, [535]). Defects in Treg have been shown to contribute to RA pathogenesis [49, 536]. *MIR146A* is associated with several disease related variants including a QTL relating to allergic asthma (P=1.30E-6, [537]) and several cancers including prostate (P=0.01, [538]) and thyroid (P=7.00E-6, [539] cancers. In addition several SNPs have been described in *MIR146A* with varying degrees of association with SLE [540].

Table 3.1: Diseases and disorders associated with ncRNA gene loci

Genomic variants in Treg ncRNA loci that have disease associations. The 'nsv-' prefix in accession numbers refer to dbVar entries and are copy number variations covering large regions of several Mbp. An 'RGD-' prefix refers to quantitative trait loci (QTL) from the Rat Genome Database. The 'rs-' prefix relates to single nucleotide polymorphism (SNP) entries in dbSNP. More specific details on relevant variants are given in Table 3.2.

ncRNA	*Variant accession numbers*	*Associated diseases and disorders*
CRNDE	nsv532013	Developmental delay and/or additional significant and other morphological phenotypes
LINC00312	nsv530211, nsv529969, nsv530239, nsv529645, nsv529996, nsv530240, nsv529998, nsv530001, nsv529957, nsv529958, RGD:1298511	Developmental delay and/or other significant developmental or morphological phenotypes, rheumatoid arthritis
PTTG3P	nsv529749, nsv532267, nsv532269, nsv532316, nsv532317, nsv53291, RGD:1331637	Seizure, developmental delay and/or other significant developmental or morphological phenotypes, cleft palate, periauricular pit , chronic obstructive pulmonary disease
LOC286442	nsv532481, nsv529062, nsv529062, nsv529317, nsv531694, nsv529586, nsv531684, RGD:1300008	Learning difficulties, global developmental delay, developmental delay and/or other significant developmental or morphological phenotypes, decreased liver function, delayed speech and language development, ventricular septal defect
MIR146A	rs2910164, rs2961920, rs73318382, rs13157399, rs57095329, rs6864584	Several cancers, asthma, systemic lupus erythematosus
ENST00000444919	nsv532791, nsv529744, nsv532796, nsv531267, RGD:1300023	Developmental delay and/or other significant developmental or morphological phenotypes, seizure, global developmental delay, blood pressure
ENST00000415387	nsv531671, nsv531672, nsv531673, nsv529667, nsv531321, nsv531677, rs10497985, RGD:1300023	Developmental delay and/or other significant developmental or morphological phenotypes, blood pressure, age-related cognitive decline

Table 3.2: Variants with disease associations with Treg relevance at Treg ncRNA loci

Additional details of ncRNA variants associated with Treg implicated diseases.

Treg ncRNA	*Variant*	*Disease association*	*P-value*	*References*
LINC00312	RGD:1298511	Rheumatoid arthritis QTL3 (familial rheumatoid arthritis). QTL locus is over 3Mbp.	0.0139	[535]
MIR146A	rs57095329	Systemic lupus erythematosus	2.74E-8	[540]
	rs2961920, rs2910164	Several cancers including prostate and thyroid	0.01; 7.00E-06	[538, 539]
	RGD:1331665	Allergic/atopic asthma related QTL 2	0.0000013	[537]

.

3.2.8 Differential gene expression between SLE patients and healthy controls

To further characterise the relevance of these Treg ncRNA to Treg biology, their expression was compared between healthy individuals and patients with SLE, the archetypal systemic autoimmune disease. As a major role of Treg is the maintenance of peripheral tolerance and the prevention of autoimmunity, defects in Treg have been suggested as a possible etiological cause of SLE. Some differences in Treg identity have been observed and include numerical and phenotypic differences, including between naïve and memory populations, and some evidence of functional defects in *in vitro* suppression assays, but observations are often contradictory. Although understanding of the contribution of Treg behaviour to SLE is not fully understood, the disease allows an *in vivo* examination of Treg in an inflammatory setting that is relevant to their function. In addition, examining the expression of novel and possibly functional Treg upregulated transcripts may inform understanding of Treg identity in an autoimmune context.

Naïve and memory Treg and Tresp were isolated from healthy individuals and SLE patients as described in Sections 2.2.1 to 2.2.6. RNA was prepared according to Sections 2.3.1 to 2.3.3 and microarrays performed as per Section 2.3.5. The age range of healthy individuals was 23 to 54 years (median: 30; standard deviation: 12) and of SLE patients was 21 to 37 years (median: 35; standard deviation: 7). All donors were female and Caucasian. The SLE patients were selected to reflect general SLE rather than subtypes of the disease, based on clinical features, and had not

received biological therapy. Patients with active disease were chosen if possible, however the exaggerated lymphopenia in these patients often impeded the preparation of these low frequency T cell subsets for microarray analysis. Details of each donor and clinical details of SLE patients are given in Table 7.26 and Table 7.27. Array data were extracted and normalised as per Section 2.5.1. Hierarchically clustering, as described in Section 2.5.2, showed trends in gene expression seen in previously published research (Figure 3.10). This included interferon alpha inducible genes, IFI27, IFI44 and IFI44L and viral response related genes, including TRIM22, which are relevant to the immune signatures seen in SLE, as described in Section 1.1.3.4. This recapitulation of published data relevant to SLE immune pathogenesis gives confidence in the data.

Transcripts exhibiting significantly different expression between SLE patients and healthy controls were identified as described in Section 2.5.4. A comparison of all CD4 T cell subsets between SLE patients and healthy individuals identified 317 transcripts as upregulated in SLE and 504 as downregulated. The identities of these are given in Table 7.11 and Table 7.12, respectively. The number of significantly differentially expressed transcripts between SLE patients and healthy controls, and their distribution between in Treg and Tresp subsets are represented in Figure 3.11

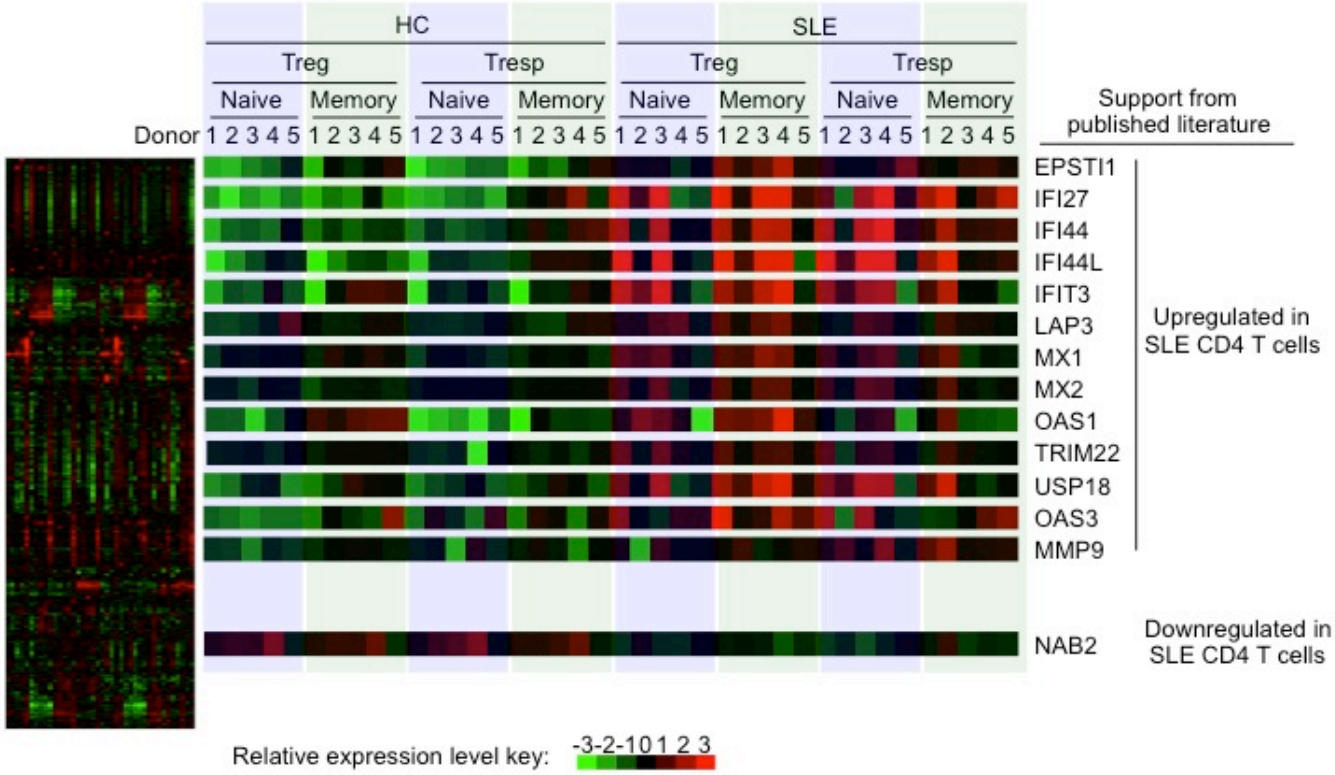


Figure 3.10: Hierarchical clustering of CD4 T cell subsets from SLE patients and healthy controls reflects previously published work

Relative expression levels in naïve and memory Treg and Tresp from SLE patients and healthy controls of genes with previous associations with SLE. HC: Healthy control; SLE: Systemic lupus erythematosus. EPSTI1: epithelial stromal interaction 1 [158]; IFI27: interferon, alpha-inducible protein [158, 541] IFI44: interferon-induced protein 44 [158, 542]; IFI44L: interferon-induced protein 44-like [543]; IFIT3: interferon-induced protein with tetratricopeptide repeats 3 [158]; LAP3: leucine aminopeptidase 3 [158] ; MX1: MX dynamin-like GTPase 1 [158]; MX2: MX dynamin-like GTPase 2 [158]; OAS1: 2'-5'-oligoadenylate synthetase 1 [158]; TRIM22: tripartite motif containing 22; USP18: ubiquitin specific peptidase 18; OAS3: 2'-5'-oligoadenylate synthetase 3 [158]; MMP9: matrix metallopeptidase 9 [544]; NAB2: NGFI-A binding protein 2 [158].

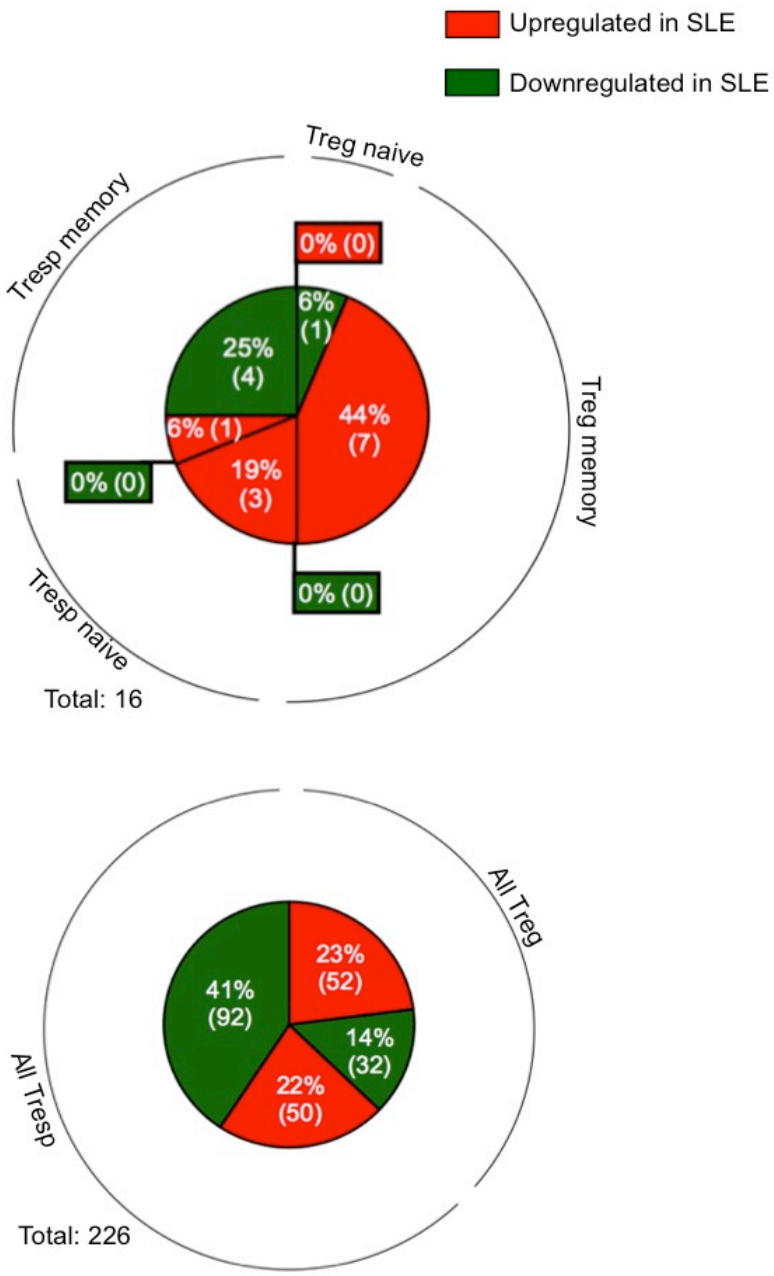


Figure 3.11: Numbers of differentially expressed transcripts in CD4 T cell subsets between healthy individuals and SLE patients
Transcripts with statistically significant differential expression between SLE patients and healthy individuals were identified by Ranksum. A: Comparison of naïve and memory Treg and Tresp subsets. B: Transcripts uniquely differentially regulated in each naïve and memory subset. The percentage of the total number of differentially expressed transcripts is given while bracketed numbers indicate the absolute number of transcripts.

As Figure 3.11 shows, when comparing naïve and memory subsets of Treg and Tresp between SLE patients and healthy individuals, few statistically significant gene expression differences were found. Only 16 genes were differentially regulated in total with 1 being seen in the naïve Treg comparison (Table 7.15 and Table 7.16), 7 in the memory Treg comparison (Table 7.17 and Table 7.18), 3 in the naïve Tresp comparison (Table 7.21 and Table 7.22) and 5 in the memory Tresp comparison (Table 7.23 and Table 7.24). The identities of these genes are detailed in Section 7.1.2. When comparing total Treg, 52 transcripts were upregulated and 32 were downregulated in SLE (Table 7.13 and Table 7.14) while in Tresp 50 genes were upregulated and 92 downregulated (Table 7.19 and Table 7.20). The identities of these genes are detailed in Section 7.1.2.

To interrogate gene expression differences relating to global changes in transcriptional regulation, transcription factors were identified according to Section 2.5.8. The differential expression of these in Treg subsets is shown in Figure 3.12. Of the 92 transcripts that were differential expressed between the Treg of SLE patients and healthy individuals, 9 were transcription factors. Three were upregulated in SLE in total Treg (MAFB, FOSL2 and NR4A3) while 5 were downregulated (HOMEZ, ZFP14, ZNF175, ZNF232, ZN404). MAFB (V-maf musculoaponeurotic fibrosarcoma oncogene homolog B) was also upregulated in the memory subset of Treg. Although expression of MAFB has been reported in T cells before [545], literature on its expression is predominantly restricted to macrophage differentiation where its role is not comprehensively understood but has been observed to increase upon stimulation [546]. Of

these transcription factors, one has a known role in Treg function: NR4A3 (nuclear receptor subfamily 4, group A, member 3), also known as NOR1. NR4A3, along with NR4A1 and NR4A2, has recently been shown to have an essential role in Treg development [547]. NR4A family transcription factor knockout mice have fewer Treg and develop autoimmunity and ChIP-Seq shows all members of the family bind the FOXP3 [547].

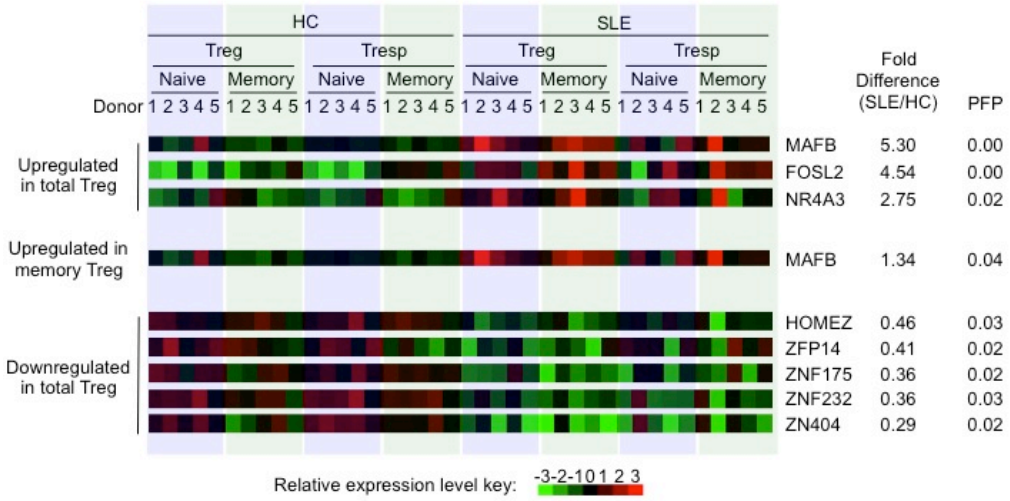


Figure 3.12: Differential expression of transcription factors in Treg between SLE patients and healthy individuals

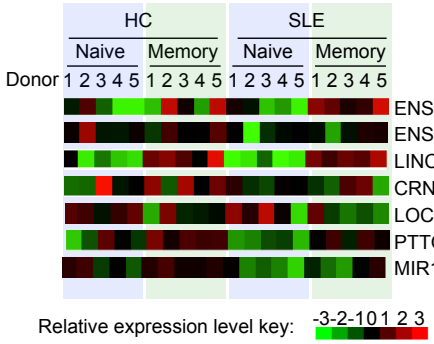
Differentially expressed transcription factors in Treg and Tresp between SLE patients and healthy individuals were identified. Their relative expression levels are shown here across 5 biological replicates with corresponding fold difference between SLE and healthy controls (HC) and the significance level. PFP: predicted false positive.

3.2.9 Differential expression of Treg ncRNAs between SLE patients and healthy individuals

I considered whether the ncRNAs preferentially expressed in Treg were deregulated in SLE patients. Figure 3.13 compares their expression levels and gives fold differences between SLE and healthy controls alongside level of significance.

None of the 7 Treg ncRNAs were statistically differentially expressed between SLE patients and healthy individuals when comparing each T cell subset individually. As an alternative approach to examining their deregulation, I compared SLE patients and healthy individuals in terms of fold differences in expression between Treg and Tresp (Figure 3.14.A) and naïve and memory Treg (Figure 3.14.B). This allowed me to highlight any trends that may tend towards significance but fall below the statistical cut-off. As expected, the known Treg relevant genes FOXP3, CTLA4, IKZF2 (Helios) and IL2RA (CD25) were more highly expressed in total Treg compared to Tresp for both SLE patients and healthy controls (Figure 3.14.A). The fold difference of these genes between Treg and Tresp in SLE was directly proportional to that in healthy individuals. This lack of difference between SLE patients and healthy controls is reflective of the published body of research that has not consistently seen differences in SLE in such Treg specific genes in the total Treg population (Section 1.1.3.6). When examining these genes between naïve and memory Treg subsets, Figure 3.14.B shows that expression levels are similar between subsets in both SLE and HC. Genes that are known to be differentially expressed between these subsets (AHR, RORC, PECAM1, Figure 3.14.A)

show their expected trends and these are similar between SLE and HC. Regarding Treg upregulated ncRNAs, subtle differences were however seen: LOC286442 and PTTG3P had a higher fold difference between total Treg and Tresp in healthy individuals compared to SLE patients, implying they are decreased in SLE Treg and/or increased in SLE Tresp (Figure 3.14.A). In addition, CRNDE showed the greatest difference in fold difference of expression in naïve Treg compared to memory Treg between SLE patients (median $\log_2$ ratio: 0.6) and healthy individuals (median $\log_2$ ratio: -1.6) (Figure 3.14.B)



	Total Treg		Naïve Treg		Memory Treg	
	Fold Difference (SLE/HC)	PFP	Fold Difference (SLE/HC)	PFP	Fold Difference (SLE/HC)	PFP
ENST00000415387	2.04	1.41	0.64	0.91	1.97	2.17
ENST00000444919	0.87	0.57	0.85	0.48	0.88	0.68
LINC00312	0.85	0.76	0.73	0.58	1.01	1.01
CRNDE	0.82	0.58	1.19	0.71	0.27	0.47
LOC286442	0.68	0.79	0.80	0.83	0.65	0.75
PTTG3P	0.49	0.62	0.54	0.56	0.65	0.68
MIR146A	0.35	0.46	0.35	0.43	0.56	0.49

Figure 3.13: Comparison of Treg ncRNA expression between SLE patients and healthy individuals.

Left: Heatmap showing the expression of Treg ncRNAs in naïve and memory Treg and Tresp from SLE patients and healthy controls. Right: Fold difference and predicted false positive rate (PFP) between SLE patients and healthy controls. HC: healthy control.

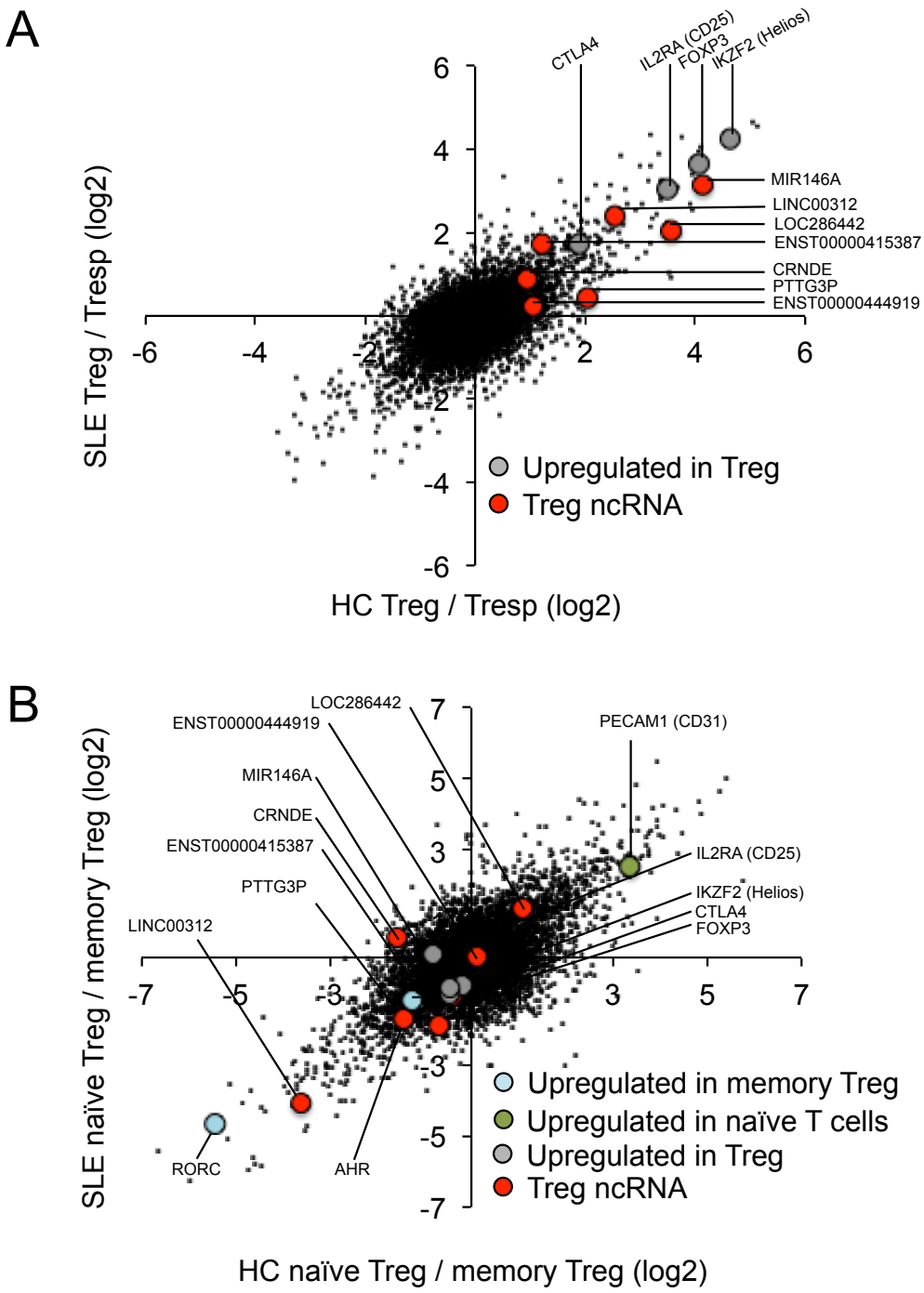


Figure 3.14: Comparison between healthy individuals and SLE patients of expression of Treg ncRNA between T cell subsets

Scatter plots of log2 fold difference in expression between T cell subsets in SLE patients and healthy controls. A: Treg compared to Tresp. B: Naïve Treg compared to memory Treg. Treg ncRNAs are highlighted in red. Genes known to be upregulated in Treg are in grey and those known to be upregulated in naïve T cells are in green. Global gene expression (i.e. all genes on the array) is shown in black.

3.2.10 Differences in histone modifications at autoimmune responsive genes

To further identify differences in Treg between SLE patients and healthy individuals at the level of transcription regulation, histone modification patterns were compared. Differences at this level would highlight mechanisms that influence the Treg transcriptional program in a physiologically relevant autoimmune setting.

The number of cells required for ChIP-Seq precluded the use of this technique for identifying histone modifications in SLE patients. SLE patients are lymphopenic and, due to the heterogeneity of the disease as well as biological variation in general, pooling of patient samples to increase material available would significantly reduce the power to identify SLE versus healthy differences. To adapt to this challenge I investigated histone modifications in healthy donors at genes that were differentially regulated between healthy donors and SLE patients. I hypothesised that genes that change in expression in an inflammatory autoimmune context would exhibit differences in histone modifications compared to genes with unchanged expression and that these differences in histone modification may relate to their responsiveness to autoimmune inflammation.

As detailed in Section 3.2.8, differential gene expression between SLE patients and healthy controls was determined through microarrays and Ranksum analysis. Histone modification ChIP-Seq data was obtained from a publically available source (Table 2.8). The lists of genes up- and downregulated in SLE Treg were compared to a background control list (all

genes on the array) for differences in profiles of histone modifications: H3K4me3 (a mark of transcriptional initiation), H3K36me3 (elongation), H3K27me3 (poised repression), H3K4me1 (enhancer function), H3K9me3 (heterochromatic repression) and whole cell extract (background enrichment). Figure 3.15 illustrates the average enrichment profile of each histone modification across differentially expressed genes.

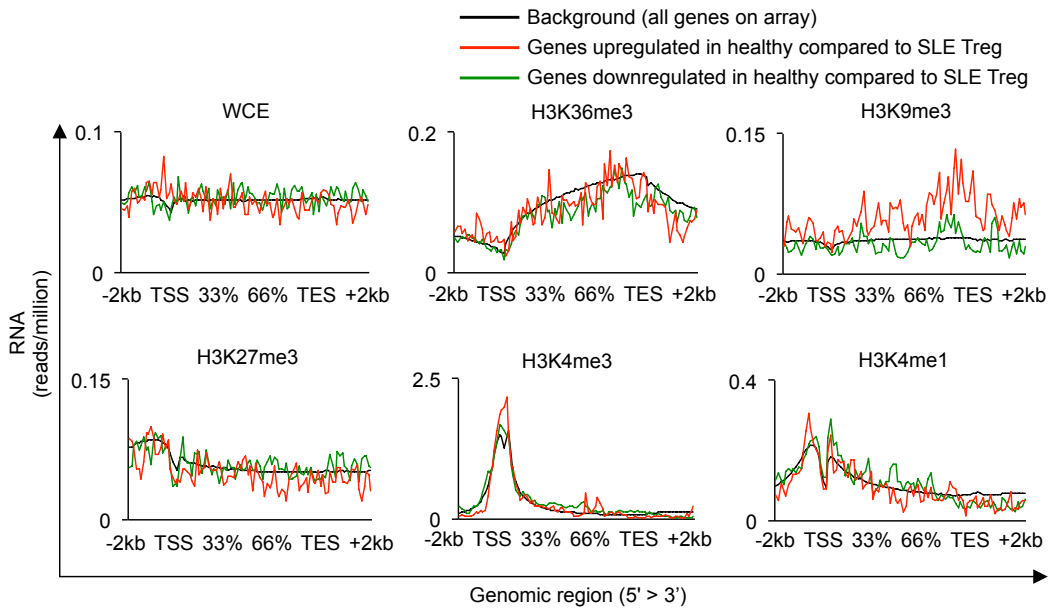


Figure 3.15: Average enrichment profile of histone modifications at genes differentially regulated in SLE Treg

Average enrichment profiles for several histone modifications showing the average number of reads mapping across the length of genes from 2kb upstream to 2kb downstream. WCE: whole cell extract. TSS: transcription start site; TES: transcription end site.

The profile of H3K9me3 was notable as it differed with genes upregulated in HC Treg compared to the background gene list. In comparison, other histone modifications showed comparable enrichment in all genes similar to the background enrichment (WCE). To determine the statistical significance of this enrichment, the cumulative frequency distribution of H3K9me3 enrichment was compared between differentially expressed genes. In Treg cells (Figure 3.16.A), genes upregulated in the Treg of healthy individuals (24 genes) had a maximum difference in cumulative distribution (D) of 0.2296 compared to a random selection of genes. Although this difference was not statistically significant (P: 0.146), it was larger and more significant than the same comparison in whole cell extract (D: 0.1939; P 0.308). In Tresp cells (Figure 3.16.B), genes upregulated in Tresp of healthy individuals compared to SLE patients (73 genes) had a maximum difference in distribution of 0.2210 (P: 0.114) while in whole cell extract this comparison was 0.1439 (P: 0.109). This demonstrated a trend towards enrichment of H3K9me3 at genes differentially expressed between SLE and HC in both Treg and Tresp. As the level of significance may have been limited by the low numbers of differentially expressed genes, and as this enrichment appeared in both Treg and Tresp, I also compared genes that were upregulated in total CD4 T cells (504 genes). This comparison is given with Treg (Figure 3.16. C) and Tresp (Figure 3.16. D) ChIP-Seq data. In Treg, the maximum difference in distribution was 0.0925, which was statistically significant (P: 0.006) and larger than in WCE (D: 0.0; P: 1.0). In Tresp, the maximum difference in distribution also statistically significant was (D: 0.1188, P: <10-3, compared to WCE: D: 0.1142; P: <10-3), showing that genes that are upregulated in total CD4 T

cells compared to SLE are enrichment for H3K9me3 in both Treg and Tresp.

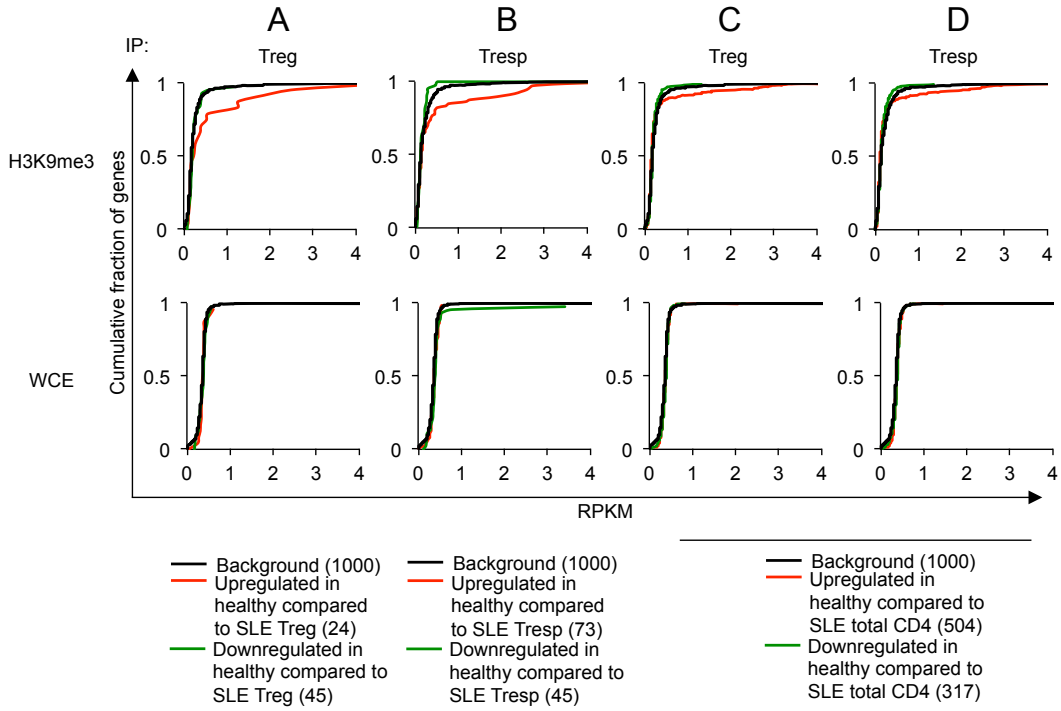


Figure 3.16: Cumulative frequency distributions of H3K9me3 enrichment at SLE deregulated genes

Cumulative frequency distributions showing the difference in H3K9me3 enrichment and WCE at genes up and downregulated in healthy controls compared to SLE patients and at a control set of randomly selected genes. A: IP performed in Treg. For genes upregulated in Treg, H3K9me3 D=0.2296 (P=0.146) and WCE D=0.0194 (P 0.308). B: IP performed in Tresp. For genes upregulated in Tresp H3K9me3 D=0.2210 (P: 0.114) and WCE D=0.0144 (P: 0.109). C: IP performed in Treg. For genes upregulated in total CD4, H3K9me3 D=0.0925 (P: 0.006) and WCE D=0.0 (P=1.0). D: IP performed in Tresp. For genes upregulated in total CD4, H3K9me3 D=0.1188 ($P=<10^{-3}$), WCE D=0.1142 (P: 0.000). IP: Immunoprecipitation; WCE: whole cell extract. RPKM: reads per killobase per million mapped reads. Brackets detail the number of genes in each list.

3.3 Discussion

Here I have identified for the first time ncRNAs with roles potentially specific to Treg biology. This identification of lincRNAs differentially expressed between Treg and Tresp is consistent with data from other labs showing that this class of ncRNAs is often cell type specific.

I identified 22 ncRNAs in total as being upregulated in Treg compared to Tresp (Figure 3.3) with differential expression of 7 being validated by qPCR in 3 additional donors. Of the other 15 Treg ncRNAs that were detected by microarray but not validated by qPCR, several showed the expected trend of increased expression in Treg but data did not meet the significance threshold of 0.05. This lower statistical significance was influenced by the number of replicates (3), the magnitude of difference in expression between Treg and Tresp and donor variation and so with an increased number of replicates the significance for differential expression of these ncRNAs may increase. Others were problematic to assay by qPCR due to their high repetitive content, a feature that has been associated with lincRNAs [403]. In Chapter 5 I have chosen to focus on the ncRNAs validated here as their more robust differential expression in Treg supports their investigation in downstream functional assays. In addition to qPCR validation I have used publically available ChIP-Seq data on enrichment of histone modifications (transcriptional initiation: H3K4me3; and H3K36me3: elongation) and FOXP3 to support their conversely regulated expression in Treg and Tresp (Figure 3.5). Histone modification data also allowed interpretation of their biological roles (H3K4me1 is a marker of enhancer RNA transcription) and FOXP3 occupancy suggested

expression of these ncRNAs might be regulated by this Treg-specifying factor. The identity of these transcripts was further probed through the genomic structure of ncRNA genes (Section 3.2.5), correlation of their expression with neighbouring genomic genes with known biological roles (Section 3.2.6) and presence of annotated disease–associated genomic variants (Section 3.2.7). The potential relevance of these Treg ncRNAs for defects in Treg function was explored through their expression in SLE, in which Treg have been shown to have phenotypic differences including in naïve and memory subsets (Section 3.2.8).

The identities of the qPCR validated Treg ncRNAs are now summarised and discussed in relation to previously published data and the data presented here:

3.3.1 MIR146a

The detection of upregulated miR-146a (ENST00000517927) in Treg is not novel but this recapitulation of previously published work [520, 521] gives confidence in the microarray data. Disease associated variants are also well documented in MIR146A compared to the other ncRNAs (Table 3.1) and include associations with Treg related diseases such as several cancers [538, 539], asthma [537] and SLE [540] (Table 3.2). Mir-146a was detected here in its pri-miRNA form, which has not previously been shown in Treg and thus demonstrates that mir-146a abundance is regulated at the level of transcription. MIR146A had a high fold difference in expression between Treg and Tresp by microarray (13.4 fold, Figure 3.3) and qPCR (21.8 fold, Figure 3.4) and as expected shows clear enrichment of H3K4me3 and H3K36me3 in Treg compared to Tresp (Figure 3.5). In addition, there is clear enrichment of FOXP3 across the

MIR146A locus (Figure 3.6). Interestingly, MIR146A shows a strong positive correlation with its genomic neighbour PTTG1 (Pearson correlation coefficient: 0.78, Figure 3.9), which is encoded in an antisense divergent orientation (Figure 3.8) and may therefore share upstream regulatory regions.

3.3.2 LINC00312

LINC00312 showed a fold enrichment of 9.4 by array (Figure 3.3) and 20.8 by qPCR (Figure 3.4) compared to Tresp and supporting this, preferential enrichment of H3K36me3 can be seen in Treg (Figure 3.5). In addition it is the only Treg ncRNA to be more highly expressed in memory compared to naïve Treg (Figure 3.3 and Figure 3.13.B). There is little enrichment of H3K4me3 at the gene for LINC00312 however the presence of H3K4me1 nearby (Figure 3.5) and its mono-exonic structure (Figure 3.7) suggest it may behave as an enhancer RNA. The *LINC00312* locus is found within a quantitative trait locus that correlates with the Treg-relevant autoimmune disease rheumatoid arthritis (Table 3.2) [49, 535, 536]. However this association maps to a large locus that contains multiple genes and so does not inform specifically about *LINC00312* activity. In addition, detection of LINC30012 could also be a result of expression of an unnanotated long isoform of the upstream gene *LMCD1*. This could also explain the lack of H3K4me3 at LINC00312 as well as the strong correlation in expression between *LINC00312* and *LMCD1* (Figure 3.9). The latter has been reported to increase in expression with calcineurin and NFAT activity [529], signalling molecules relevant to T cell activation that are more active in memory T cells.

3.3.3 PTTG3P

PTTG3P is a pseudogene and therefore less likely than other non-coding transcripts to have evolved to mediate a function at the transcript level. While there is FOXP3 binding near *PTTG3P* (Figure 3.6), this could relate to the presence of the antisense gene *SK3G* (Figure 3.8) and the lack of enrichment of neither H3K4me3 nor H3K4me1 (Figure 3.5) may suggest it is not an actively regulated gene in Treg.

3.3.4 ENST00000444919

Although significant, the increased expression of *ENST00000444919* in Treg compared to Tresp was low by both array (2.45 fold, Figure 3.3) and qPCR (1.4 fold, Figure 3.4) and in line with this enrichment of H3K4me3 is similar in both T cell subsets (Figure 3.5). *ENST00000444919* does show FOXP3 binding at the promoter (Figure 3.6) suggesting it is actively regulated in Treg and has a large intron to exon ratio with the presence of several repeat elements similar to other ncRNAs (Figure 3.7). However the upstream region, which may harbour regulatory elements, is shared with that of *PDK1,* which is in a divergent orientation (Figure 3.8), and the low correlation in expression between the two (Figure 3.9) may suggest this FOXP3 only regulates the *PDK1* gene.

3.3.5 ENST00000415387

ENST00000415387 is encoded at the 3' end of *IKZF2,* which encodes the Treg transcription factor Helios (Figure 3.8 [93]), and the two are strongly correlated in expression (Figure 3.9), which is consistent with ENST00000415387 having a role in Treg biology through regulating *IKZF2* expression. However due to the orientation of *ENST00000415387* with

IKZF2 the presence of an unannotated or unspliced isoform of IKZF2 could also explain the increased signal by array and qPCR. Supporting this, the TSS for *ENST00000415387* lacks a H3K4me3 peak while upstream on can be found at the TSS of *IKZF2* (Figure 3.5).

3.3.6 CRNDE

CRNDE has distinct H3K4me3 enrichment at the TSS while H3K4me1 is minimal (Figure 3.5), which is consistent with a role as a lincRNA rather than enhancer RNA. In addition, there may be some, although low, occupancy of FOXP3 at the promoter (Figure 3.6), which supports regulation of the gene in Treg. CRNDE has obtained its name from its observed upregulation in colorectal adenomas and adenocarcinomas, an early event in colorectal neoplasias [524]. CRNDE was classified with relative high confidence as a lincRNA in an evaluation of genome-wide annotations [496]. In other literature it is referred to as lincIRX5, after the protein-coding gene it neighbours [441], and linc1399 after the mouse ortholog [405]. CRNDE is also upregulated in hepatocarcinoma [548], haematological malignancies [524] and brain cancer [524]. In support of a functional role for CRNDE in cancer, genes affected by siRNA knockdown of CRNDE in fibroblasts are enriched for processes relating to cancer, cell death and cell cycle [441, 524]. Comparison of SLE patients and healthy individuals showed here that of all Treg ncRNAs, CRNDE showed a greater tendency to be decreased in memory Treg in SLE (Figure 3.14.B). This may reflect a lower proliferative status of memory Treg in SLE, which may relate to the decreased activity of this Treg subset that has been observed by others [96] (Section 1.1.3.6.1). In addition to cancer cells, CRNDE has been observed to be highly expressed in embryonic stem

cells [442] and has been associated with pluripotency, based on the observation that expression decreases with specialisation [526]. The exception is neurogenesis where expression increases compared to ESCs, although it becomes undetectable in adult neurons [526, 549].

Enforced expression of a CRNDE transcript containing a highly conserved intronic region termed genomic vertebrate conserved intron 4 (gVCIn4) promotes growth and suppresses apoptosis in fibroblasts [526]. Unspliced transcripts containing intronic regions, including gVCIn4, are downregulated by insulin and insulin-like growth factor (IGF), an effect that is inhibited by blockade of the insulin signalling pathways PI3K/Akt/mTOR and Raf/MAPK [550]. Knockdown of gVCIn4 containing transcripts with siRNA in fibroblasts affects expression of insulin related signalling molecules and downstream processes including glucose and lipid metabolism [550]. Therefore CRNDE transcripts may antagonise events downstream of insulin signalling. Knockdown of spliced isoforms of CRNDE did not have this effect and were not downregulated by insulin in fibroblasts [550]. CRNDE has been detected in T cells before although at a low level compared to other cell types such as spermatozoa [526] and consistent with a potential role in pluripotency, expression in thymic T cells decreases from the double positive (CD4+CD8+) to single positive (CD4+CD8- and CD4-CD8+) states [551]. Increased expression in of CRNDE Treg compared to Tresp has not been previously reported.

The expression of lncRNAs has been suggested as a mechanism for the genomic targeting of chromatin-modifying complexes for the regulation of transcription of specific genes through cis-regulatory mechanisms (Section

1.3.3.1. [441]). Consistent with this, the promoters for CRNDE and its neighbouring gene IRX5 are adjacent, share the same CpG island and a comparison across several tissue types shows their expression is correlated [526], suggesting that CRNDE transcription may influence IRX5 transcription. Furthermore CRNDE has been shown to physically associate with the chromatin-modifying complexes PRC2 and CoREST by native RNA immunoprecipitation [441] and knockdown of PRC2 and CRNDE show overlap in the biological processes affected [441]. Although CRNDE and IRX5 do not strongly correlate (Figure 3.9) and there is no published Treg-relevant role for IRX5, CRNDE may instead influence recruitment of these chromatin-modifying complexes to other genes. However, this evidence should be treated cautiously as the use of native RNA immunoprecipitation is limited for demonstrating a direct interaction. In addition, the theory of specific interactions of lncRNAs and PRC2 has been recently challenged by the observation that broad range of RNAs, including mRNA, have a capacity to bind PRC2 [552, 553]. The repeats seen in CRNDE that correspond with exonic sequences (Figure 3.7) support the importance of secondary structure in the transcribed product, which may be required for interaction with chromatin modifiers.

3.3.7 LOC286442

LOC286442 had the highest fold difference between Treg and Tresp of any ncRNA by array (13.9 fold, Figure 3.3) and supporting this differential expression there was clear greater enrichment of H3K4me3 and H3K36me3 at the *LOC286442* gene in Treg compared to Tresp (Figure 3.5). In addition FOXP3 binds at the promoter and in intronic regions (Figure 3.6) supporting the possibility of Treg-specific regulation of the

gene. In addition, this ncRNA is of interest as there is a higher fold difference between total Treg and Tresp in healthy individuals compared to SLE patients, implying they are decreased in SLE Treg and/or increased in SLE Tresp (Figure 3.14.A).

Based on the conclusions above, of the ncRNAs upregulated in Treg compared to Tresp, LOC286442 and CRNDE appeared to have the most potential for functional assays investigating their biological relevance in Treg and this is the subject of Chapter 5.

3.3.8 Transcriptional differences in SLE Treg

In both Treg and Tresp cells from a healthy individual H3K9me3 was enriched at genes that were increased in expression in CD4 T cells of healthy individuals compared to SLE patients (Figure 3.15, Figure 3.16). This enrichment at differentially expressed genes was significant when compared to enrichment at randomly selected genes (Figure 3.16). This was in comparison to other histone modifications (H3K4me3, H3K4me1, H3K36me3 and H3K27me3) and WCE, which did not show an obvious difference (Figure 3.15). This was unexpected as the majority of literature on H3K9me3 reports its enrichment in repressed chromatin including at peri-centromeres and at repeats and transposons [361, 362]. In these regions the presence of H3K9me3 is understood to repress deleterious transcriptional activity in part through recruitment of HP1 [364]. However a minority of H3K9me3 related literature reports its enrichment along the gene body of actively transcribed genes with Vakoc et al. 2005 first demonstrating this [363]. Vakoc et al found that tri-methylation of H3K9 was induced during activation and lost when transcription was inhibited.

This was seen in murine cell lines as well as primary T cells stimulated through CD3, including at the *Il2* gene [363]. H3K9me3 was increased along the gene body compared to the promoter [363], as is seen here (Figure 3.15), and inhibition of phosphorylation of the RNAPII CTD led to a reduction in H3K9me3 along the gene body of active genes but not at microsatellites, suggesting a difference in the role of H3K9me3 in active compared to inactive chromatin and a relationship with transcriptional elongation. Vakoc et al. also demonstrated the recruitment of the H3K9me3 associated chromatin-modifier HP1γ and an interaction between this and elongating RNAPII [363]. Others have since shown that presence of HP1γ may slow the processivity of RNAPII leading to inclusion of exons [371], therefore demonstrating a relationship between H3K9me3 and alternative splicing, which is supported by the enrichment of H3K9me3 at included exons of the *CD44* gene [371]. This suggests that the differential expression between HC and SLE may relate to regulation of elongation and/or alternative splicing. Regulation of gene expression at this point may allow response to an inflammatory environment. H3K36me3, which is also associated with elongation, did not appear to differ at differentially expressed genes, suggesting that H3K9me3 but not H3K36me3 mediated transcriptional elongation is associated with the regulation of this inflammation related set of genes.

Through examining differential expression in Treg between SLE patients and healthy controls, I identified several transcription factors of potential interest that may contribute to differential gene expression. This included MAFB (V-maf musculoaponeurotic fibrosarcoma oncogene homolog B)

and NR4A3 (Figure 3.12, Section 3.2.8). MAFB was upregulated in Treg in general as well as in the memory (CD45RA-) subset of Treg in SLE patients. A recent profile of differential expression of transcription factors between CD45RA+ and CD45RA- Treg has demonstrated that CD45RA- Treg express higher levels of transcription factors of other lineages [328]. This includes a MAF family member, c-MAF (V-maf musculoaponeurotic fibrosarcoma oncogene homolog), which has a known role in T cell differentiation in promoting both Th2 [554, 555] and Th17 [556] phenotypes and regulates transcription of targets of ETS1 transcription factor targets [557]. Given the structural similarity of the MAF family members, which bind DNA though a basic leucine zipper domain, the role of c-MAF in T cell differentiation may inform the relevance of MAFB upregulation in SLE. The transcription factor NR4A3 was also identified as being upregulated in SLE Treg. NR4A3 has been recently identified as having a role in Treg development [547]. This upregulation in SLE may reflect a role of NR4A3 in supporting the response of Treg to the inflammatory environment and may demonstrate that SLE Treg are capable of this response.

4 Identification of enhancer RNAs in Th1 cells

4.1 Introduction

4.1.1 T-bet super-enhancers

Our lab has reported that the majority of T-bet binding sites are located distally to known gene promoters [246], suggesting the regulatory action of T-bet is mediated through enhancers. This binding often occurs in domains containing five or more distinct T-bet peaks [246]. At individual gene loci, including the IFNG locus, [246], these extended T-bet domains correlate with DHS *sites* and the presence of H3K4me1, indicating an open chromatin conformation suggestive of functional enhancer elements [246]. The high density T-bet bound regions upstream of IFNG have orthologs in mouse [246]. In genome-wide analyses, genes distal to such domains are enriched for biological processes related to immune response, supporting a biological significance for these extended T-bet bound domains *[246]*. Furthermore, T-bet binding at distal elements in addition to proximal elements, as opposed to proximal elements alone, is more likely to correlate with lineage specific gene expression. Supporting the importance for distal binding of T-bet, knockout of Tbx21 (T-bet) in mouse has demonstrated that genes bound both distally and proximally by T-bet were four times more likely to be dependent on T-bet expression that those only bound proximally. Similar data were seen for Gata3 suggesting this phenomenon is representative of lineage-specifying transcription factor behaviour in general.

The extended T-bet cis-regulatory regions identified by Kanhere and colleagues resembles recently defined super-enhancers (described in

Section 1.2.2.3). To confirm this, Richard Jenner used the ROSE (rank ordering of super-enhancers) algorithm [320, 321], which stiches neighbouring enhancers together and ranks these by transcription factor binding [321], to define super-enhancers using in-house human and mouse T-bet ChIP-Seq data. The output of the ROSE algorithm for T-bet in human Th1 cells is shown in Figure 4.1 (made by Richard Jenner), which shows that rather than the amount of T-bet binding having a linear relationship with rank across all T-bet enhancers, a subset of enhancers, super-enhancers, exhibit markedly high T-bet enrichment. Using this approach, our lab has identified 374 super-enhancers and 2817 typical enhancers in human Th1 cells (data currently in peer review).

4.1.2 Recruitment of P-TEFb to super-enhancers and lineage-specific genes

While exploring the mechanism of super-enhancer function and lineage-specific gene expression, our lab has observed comparable RNAPII binding in the promoter region of Th1-specific genes in both Th1 and Th2 cells (Figure 4.2.A, Arnulf Hertweck and Richard Jenner), implying that recruitment of RNAPII is not regulated in a lineage-specific manner. However, differential RNAPII binding is observed along gene bodies (Figure 4.2.A, Arnulf Hertweck and Richard Jenner), suggesting that RNAPII processivity through transcriptional elongation is regulated in a more lineage-specific manner than transcriptional initiation. To explore the control of this elongation, ChIP-Seq was performed to assess the lineage specificity of binding of the elongation factor P-TEFb (Figure 4.2**,** Richard Jenner and Arnulf Hertweck). Upon stimulation, RNAPII and P-TEFb show greater lineage-specific enrichment at genes associated with super-

enhancers and compared to typical enhancers. This showed that transcriptional elongation through P-TEFb recruitment correlated with the increased T-bet binding that defines super-enhancers. Additionally, extensive P-TEFb binding was observed at the super-enhancers themselves in addition to genes (Figure 4.2.B).

Given that eRNAs have been identified to be transcribed from enhancers and that P-TEFb is bound at these sites, these data suggest that eRNAs are also transcribed from T-bet super-enhancers. Furthermore, knockdown of eRNAs transcribed from other enhancers in other cell types has been reported to repress expression of neighbouring genes (discussed in Section 1.3.2.4). Linking these observations, P-TEFb has also recently been shown to be required for eRNA transcription [385].

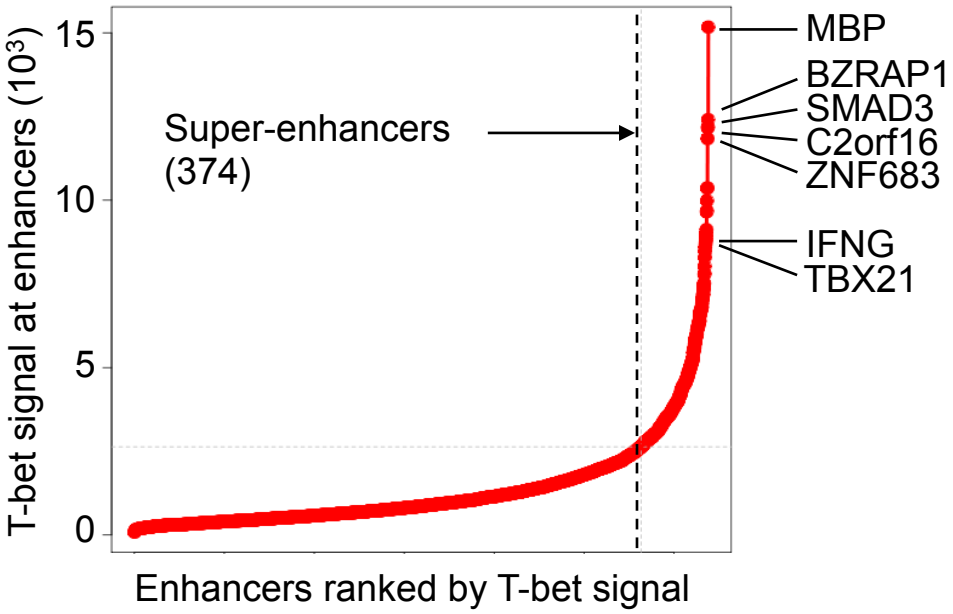


Figure 4.1: Identification of super-enhancers

T-bet bound enhancers ranked by signal and the relationship between rank and signal examined to identify super-enhancers. Super-enhancers were distinguished from typical enhancers by their markedly increased enrichment of T-bet (performed by Richard Jenner). The closest genes to the five highest ranking super-enhancers are indicated on the left. The super-enhancers closest to the Th1 signature genes IFNG and TBX21 are also indicated.

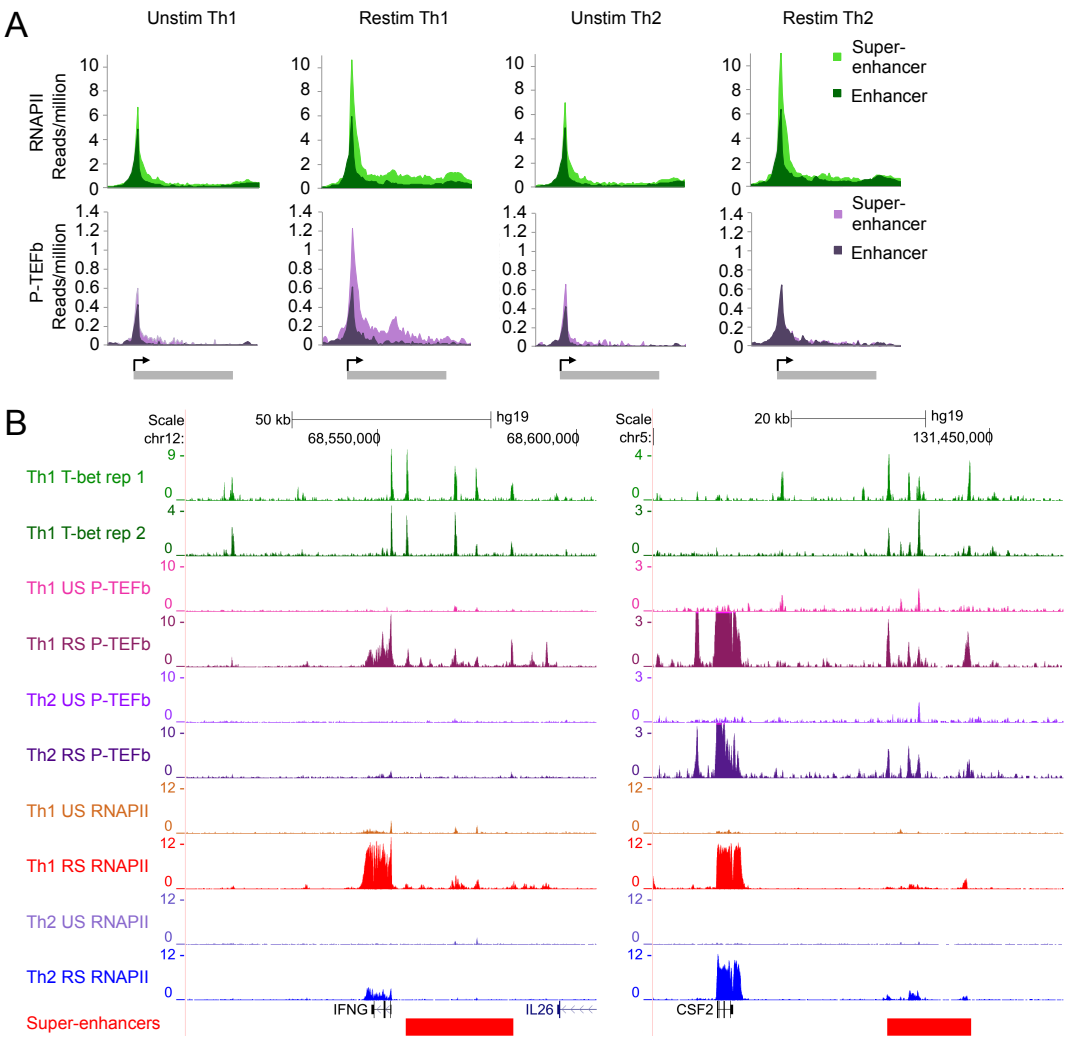


Figure 4.2: T-bet occupancy at lineage-specific genes and super-enhancers
A: Average number of ChIP-seq reads for RNA pol II and P-TEFb (reads/million, input subtracted) in unstimulated (US) and restimulated (RS) human Th1 and Th2 cells. All genes are bound by RNA pol II in at least one condition and divided into those associated with a super-enhancer (n=231) or a typical enhancer (n=1307). B. ChIP-seq binding profiles (reads/million) for T-bet, RNA pol II and P-TEFb in unstimulated and restimulated human Th1 and Th2 cells at example Th1 genes.

4.1.3 Aims

1. To examine the transcriptional output of T-bet super-enhancers genome-wide and at specific genes.
2. To characterise super-enhancer transcription in terms of features previously seen at enhancers including polyadenylation and bidirectionality.
3. To examine the relationship between T-bet and P-TEFb binding with super-enhancer eRNA transcription.

4.2 Results

4.2.1 Genome-wide strand-specific RNA-seq of Th1 and Th2 cells

I aimed to perform RNA-Seq on human Th1 and Th2 cells and naïve precursors as this would allow *de novo* identification of previously unannotated transcripts that may emanate from enhancer regions. As published research has shown that enhancer RNA may lack polyadenylation, I intended to compare enrichment of RNA from T-bet super-enhancers in RNA-Seq libraries generated from polyadenylated (polyA) RNA and total cellular RNA, which would include both polyA and non-polyA RNA. Another reported characteristic of enhancer RNA is bidirectionality i.e. transcription from both DNA strands emanating from a common start site, and in order to detect this strand specific sequencing was required. As RNA sequencing had not been previously performed in our lab, I optimised several library generation conditions.

Fragmentation of RNA prior to library generation (Section 2.3.6.2) is required due to current limitations of sequencing technologies, which can only read relatively short sequences efficiently. However, over fragmentation leads to the generation of reads that are too short and the unnecessary sequencing of PCR adapters and primers at the expense of information on the intervening insert of interest. The optimal library fragment size for the Illumina HiSeq is 180 to 200bp. In addition, homogeneity of fragment size reduces sequencing bias. As Figure 4.3 shows, two incubation times for chemical fragmentation were compared. Four minutes was chosen because fragment size was shorter and more

uniform.

In addition, the number of cycles during PCR library generation required optimisation. Too few would result in reduced sensitivity to low abundance reads while too many would lead to excessive duplication and decreased complexity. I compared generation of PCR libraries using the minimum and maximum number of cycles (12 and 15, respectively). Ribosome depleted RNA from 1µg total RNA, a yield of around 5ng, was used for this test as this reflected the lower limit of RNA available for use. Therefore this also allowed a test for the ability to prepare sequencing libraries from low quantities of RNA. As Figure 4.4 shows, 15 cycles produced a clear smear between 150 and 350 bp, which was the expected size of the library and demonstrated the ability to generate sequencing libraries from this limited amount of starting material. Twelve cycles, also produced this smear although less distinct as expected. Both conditions produced a clear resolution between library products and the primer dimer band at ~100bp. In order to maximise sensitivity of low abundance reads from the low amount of starting material, 15 cycles was ultimately used for library generation from samples of interest.

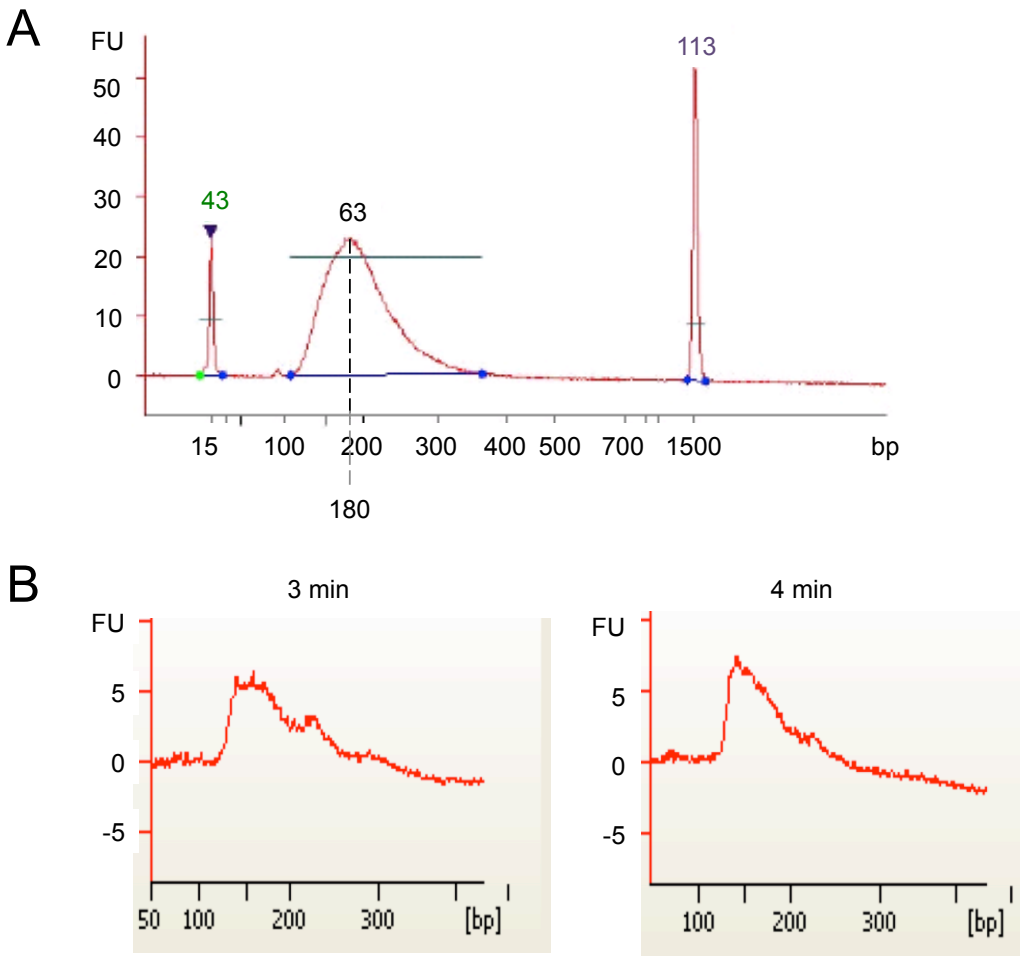


Figure 4.3: RNA fragmentation optimisation prior to RNA-Seq library generation
Bioanalyzer electropherograms showing a comparison of library size under different incubation times for chemical fragmentation of total RNA. A: Illumina directional mRNA-Seq electropherogram showing the optimal fragment length (180-200bp) for an Illumina HiSeq sequencer. Upper and lower peaks (43 and 113 bp) are size standards. B: Comparison of incubation time showing increased uniformity in fragment size with increased incubation time.

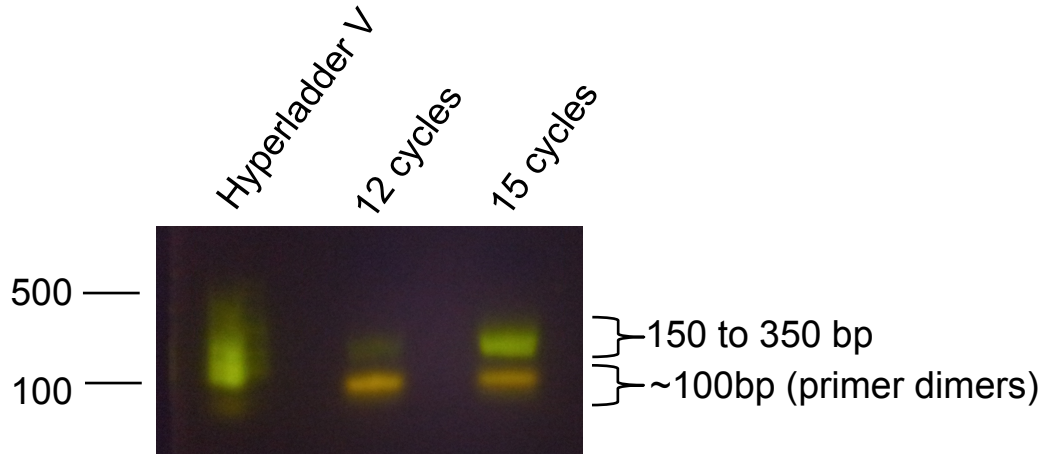


Figure 4.4: PCR library generation test

Due to the low amount of material available, test PCR libraries were generated from 1µg of ribosome depleted total RNA from CD4 T cells. A comparison was made of the lower (12) and upper (15) recommended cycling repetitions.

I performed a preliminary analysis of Th1 and Th2 polarisation in three donors by microarray before selecting the most appropriately skewed samples for RNA-Seq. Assessment of polarisation was based on expression of the Th1-specific genes TBX21 and IFNG and the Th2-specific genes GATA3, IL13, IL4 and IL5 in naïve precursors and Th1 and Th2 cells harvested at several time points during T cell polarisation cultures (cultures performed by Ian Jackson). Cells were harvested at days 3, 7 and 13 with a comparison of unstimulated and restimulated cells at day 13 As shown in Figure 4.5, for all three donors, IFNG and TBX21 were upregulated in Th1 polarising conditions at all time points compared to naïve precursors. Th2 polarising conditions also led to consistent upregulation of GATA3 compared to both naïve precursors and Th1 cells. IL4 expression in Th2 cells was greater at d13 upon restimulation but was comparable to Th1 cells. This less robust Th2 polarisation was also reflected in expression of IL13, which was not Th2 specific in any of the three donors. Donor 2 however exhibited the most lineage appropriate expression of IL5 and was selected for RNA-Seq experiments. Donor 3 was discounted due to the increased expression of IFNG seen in restimulated cells from day 13 and donor 1 taken forward as a biological replicate instead. RNA-Seq libraries from total and polyadenylated RNA from donors 1 and 2 with day 13 unstimulated and restimulated Th1 and Th2 cells were prepared and sequenced as described in Section 2.3.6.

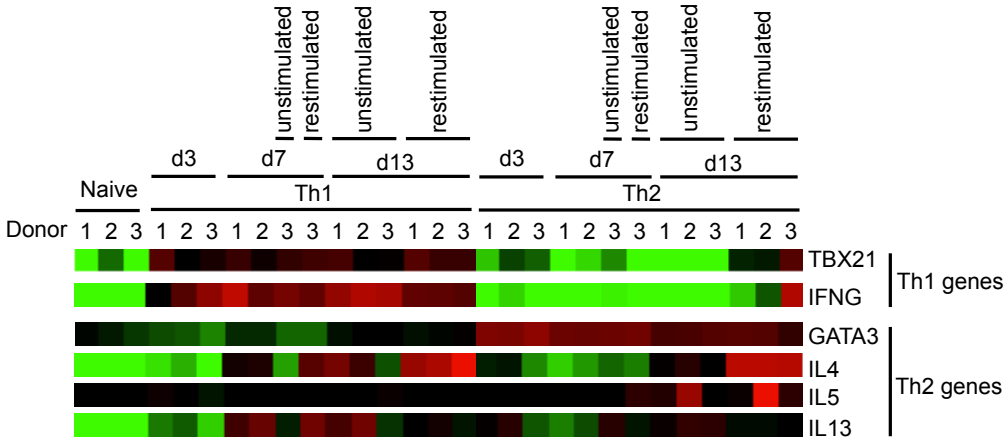


Figure 4.5: Assessment of Th1/2 polarisation for RNA sequencing by microarray. Naïve precursors from 3 biological replicates were sorted and cultured under Th1 and Th2 polarising conditions (by Ian Jackson) and lineage specific gene expression assessed by microarray. Cells were harvested at days 3, 7 and 13. Cells were harvested with and without additional restimulation for donor 3 at day 7 and all donors at day 13.

4.2.2 Super-enhancer RNA production is Th1 specific and increased upon stimulation

T-bet super-enhancers were distinguished from typical enhancers (Richard Jenner) using T-bet ChIP-Seq data from two replicate ChIP-Seq experiments (performed by Richard Jenner and Arnulf Hertweck) as described above and in Section 2.5.10. As production of RNA from enhancers has been associated with lineage specific gene expression [385, 387, 421, 425, 428, 429, 558-560], I first examined T-bet super-enhancer for their lineage specific transcription. Figure 4.6.A illustrates the average profile of RNA production at T-bet bound sites in super-enhancers in Th1 and Th2. Donor 1 does not exhibit differential RNA production from these sites. This lack of lineage specific transcription at these sites is reflected in the microarray data, which showed poorer polarisation, particularly in respect to IL4 expression (Figure 4.5). Donor 2, which was well polarised according to transcription factor and cytokine expression as measure by microarray (Figure 4.5), exhibited a clear greater production of RNA at super-enhancers in Th1 cells. Figure 4.6.B shows the cumulative distribution of RNA production from these sites, which has a maximum difference between Th1 and Th2 cells of 0.2082 from donor 2, which is statistically significant ($P<10^{-3}$).

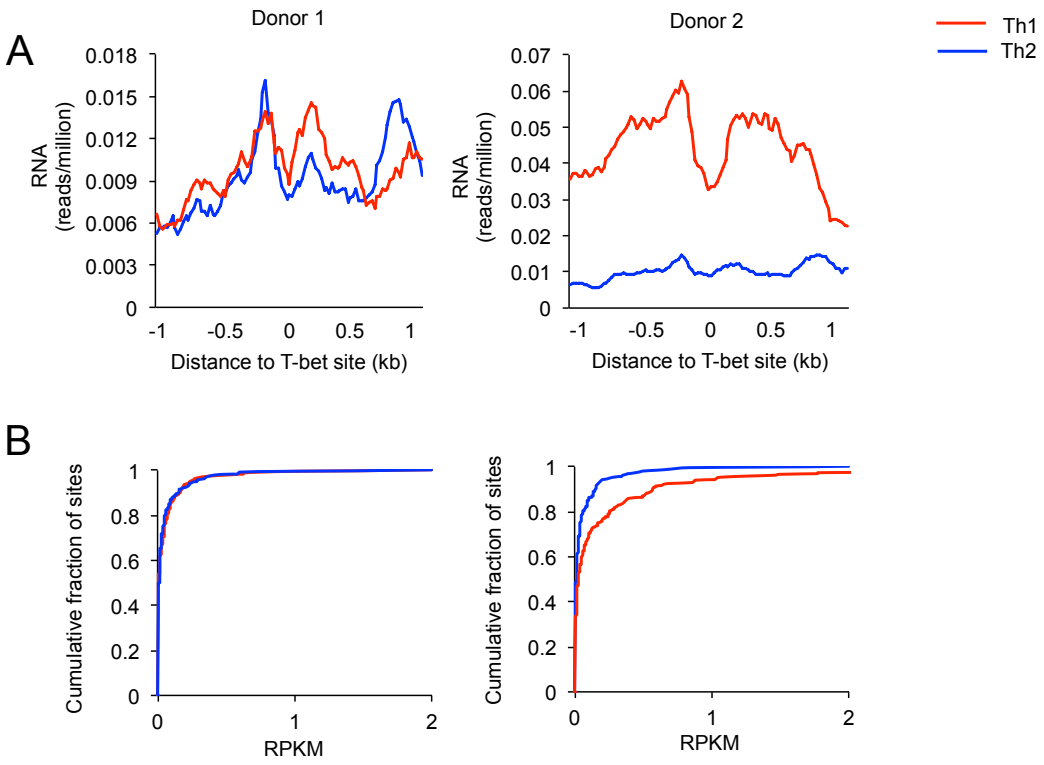


Figure 4.6: T-bet super-enhancers show lineage specific RNA production
Comparison of RNA production in T-bet super-enhancers in Th1 and Th2 cells. A: Average profile of RNA (reads/million mapped reads) in super-enhancers centered on T-bet site locations in Th1 cells. B: Cumulative frequency distributions for the amount of RNA produced from these sites in Th1 and Th2 cells. KS tests on distributions showed no statistical difference for donor 1 (D: 0.1115; P: 0.065) but donor 2 showed a maximum difference of 0.2082 ($P<10^{-3}$). RPKM: reads per kilobase per million mapped reads.

Stimulation of CD3 and CD8 is required for T cells to become activated and elicit a pro-inflammatory response. This stimulation leads to remodelling of lineage-specific and effector loci, such as *IFNG* [228, 337, 338]. To see if this stimulation affects the transcriptional activity of super-enhancers, I compared RNA sequencing data from Th1-polarised unstimulated and restimulated cells. Similar to Figure 4.6, donor 1 did not show this trend, which reflects the poor polarisation seen in this donor. As Figure 4.7 shows, restimulated cells from donor 2 exhibited increased RNA production at super-enhancers than unstimulated cells. This is seen in the average profile around T-bet sites Figure 4.7.A and the maximum difference in cumulative distribution of RNA between unstimulated and restimulated cells, 0.1190 (Figure 4.7.B), is statistically significant (P=0.041, KS test).

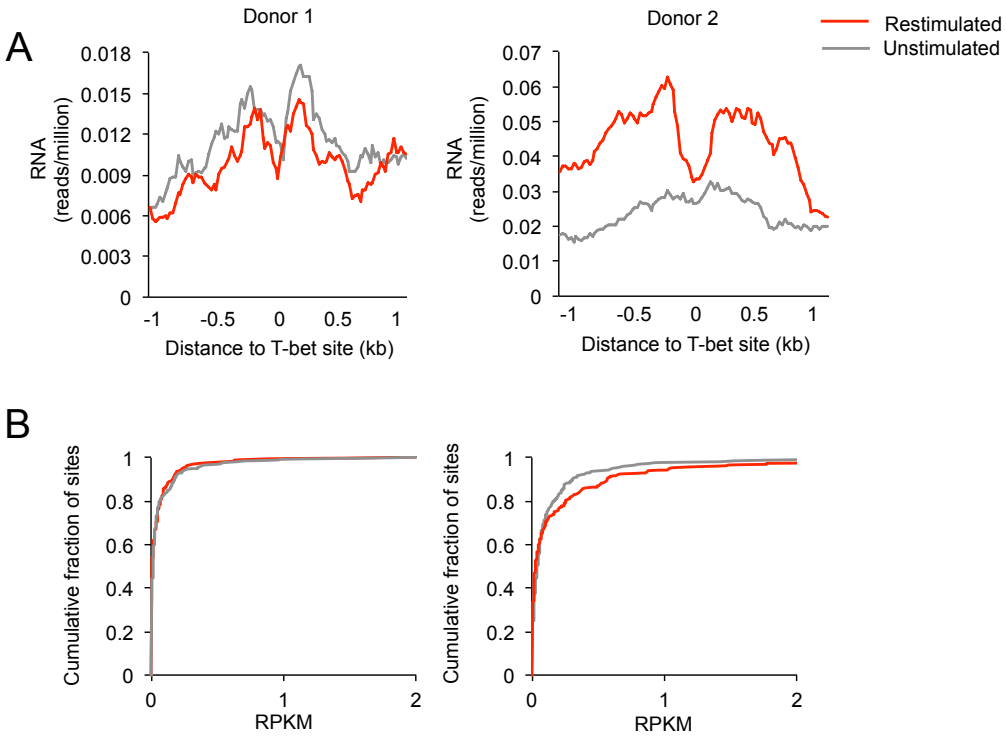


Figure 4.7: RNA production at super-enhancers increases upon stimulation
Comparison of RNA production in T-bet super-enhancers in Th1 cells before and after stimulation. A: Average profile of RNA (reads/million mapped reads) centered on T-bet sites. B: Cumulative frequency distributions for the amount of RNA produced from these sites before and after stimulation. KS tests on distributions showed no statistical difference for donor 1 (D: 0.0967; P: 0.153) but donor 2 showed a maximum difference of 0.1190 (P: 0.041). RPKM: reads per kilobase per million mapped reads.

4.2.3 Characterisation of super-enhancer eRNAs at specific genes

RNA-Sequencing data was processed for visual representation in the UCSC genome browser as described in Section 2.5.9. This allowed visual inspection of specific super-enhancer loci at Th1 relevant genes in comparison with ChIP-Seq data. Figure 4.8 shows four super-enhancers, which are proximal to genes expressed in Th1 cells. This includes the Th1 specific cytokine *IFNG* and Th1 lineage-specifying transcription factor *TBX21* (T-bet) in addition to the cytotoxic serine protease *GZMB* (Gramzyme B), which has been associated with Th1 cells [561] and the pro-protein protease *FURIN*, which is preferentially expressed on Th1 cells [562]. At each super-enhancer the presence of transcripts can be seen in the total RNA fraction from Th1 restimulated cells. In accordance with average profiles (panel A from Figure 4.6 and Figure 4.7) and cumulative frequency distributions (panel B from Figure 4.6 and Figure 4.7) of RNA production around T-bet sites, these transcripts are specific to Th1 cells and increase upon restimulation. As the ChIP-Seq data shows, this eRNA production is coincident T-bet binding with the most abundant RNA-Seq reads aligning with T-bet peaks. A role for T-bet in this eRNA production is also suggested by the presence of RNAPII at T-bet sites. In addition, RNAPII binding at these eRNA loci increases with stimulation and P-TEFb binding also follows this trend. In addition, as the presence of H3K4me3 at enhancer RNA loci is debated [289, 293, 344, 356], I examined ChIP-Seq data available from our lab, confirming that these eRNA loci are marked by H3K4me3 (Figure 4.8).

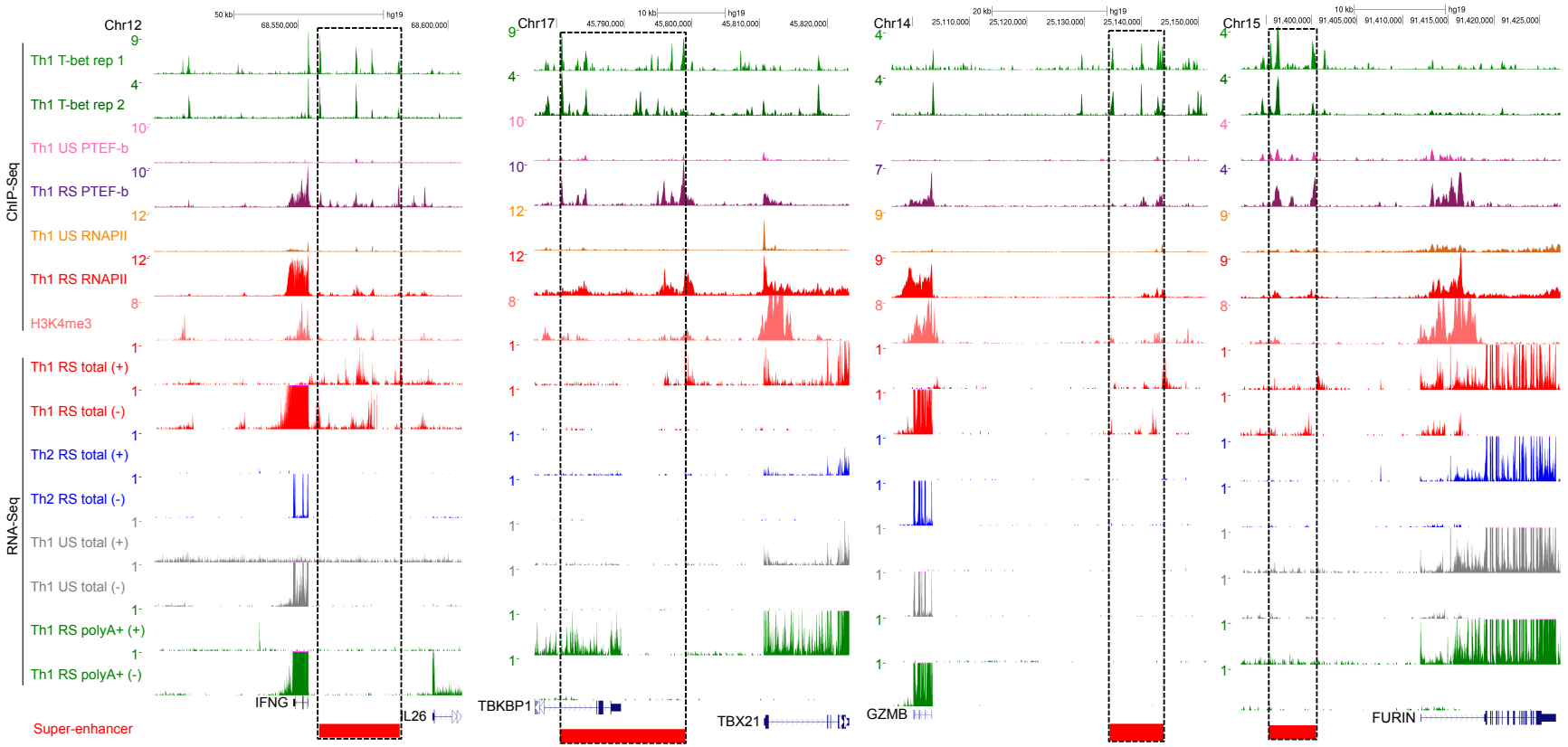


Figure 4.8: Characterisation of super-enhancer eRNAs at specific genes

(See previous page) RNA-Seq and ChIP-Seq data at super-enhancers near several Th1 relevant genes were visualised using the UCSC genome browser. ChIP-Seq data shows T-bet binding from two replicate experiments, P-TEFb and RNAPII binding in unstimulated and restimulated cells and H3K4me3 RNA-Seq tracks show positive (+) and negative (-) strand data from Th1 restimulated total RNA (red) in comparison with RNA-Seq from Th2 (blue) and unstimulated (grey) cells and RNA from the polyadenylated fraction (green). Red bars at the bottom and dotted lines specify super-enhancer locations and genes proximal to these are annotated, From left to right are super-enhancers upstream of IFNG (IFNγ), TBX21 (T-bet. This super-enhancer also overlaps with TBKBP1 [TBK1 binding protein 1]), GZMB (granzyme B) and FURIN (furin). All data are from donor 2. US: unstimulated; RS: restimulated.

4.2.4 Super-enhancer eRNAs are not polyadenylated

Comparison with RNA-Seq data from the polyA+ fraction shows that these eRNAs are enriched in the total RNA fraction (Figure 4.8) demonstrating that they lack polyadenylation, a feature that has been reported of other enhancer RNAs [420, 421]. To examine if this was a genome-wide feature of super-enhancer transcripts I compared total and polyA+ RNA fractions in terms of the average profiles and cumulative distribution of RNA production around T-bet sites at super-enhancers. As Figure 4.9 shows, the total RNA fraction from donor 2 showed clear enrichment of RNA from enhancers compared to the polyA+ purified fraction. The maximum difference in distribution was 0.1673 (P=0.001). For donor 1, which showed poorer polarisation by microarray than donor 2 (Figure 4.5), a small enrichment was seen in total RNA (D:0.1041) which was not significant (P=0.102).

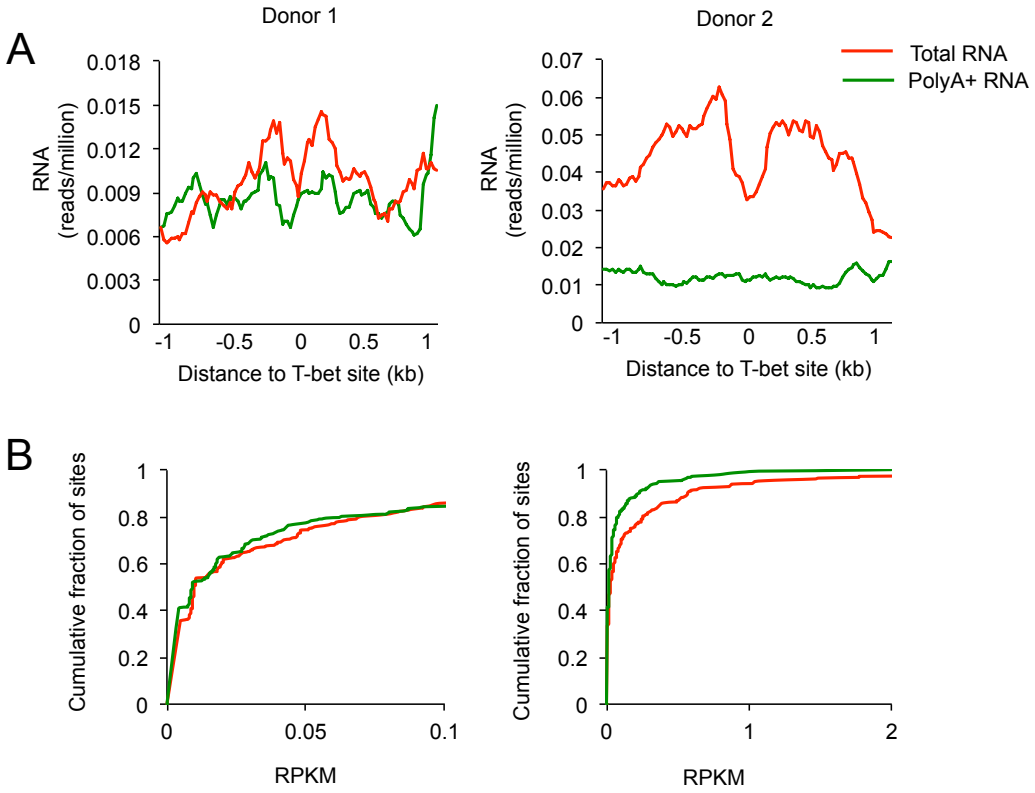


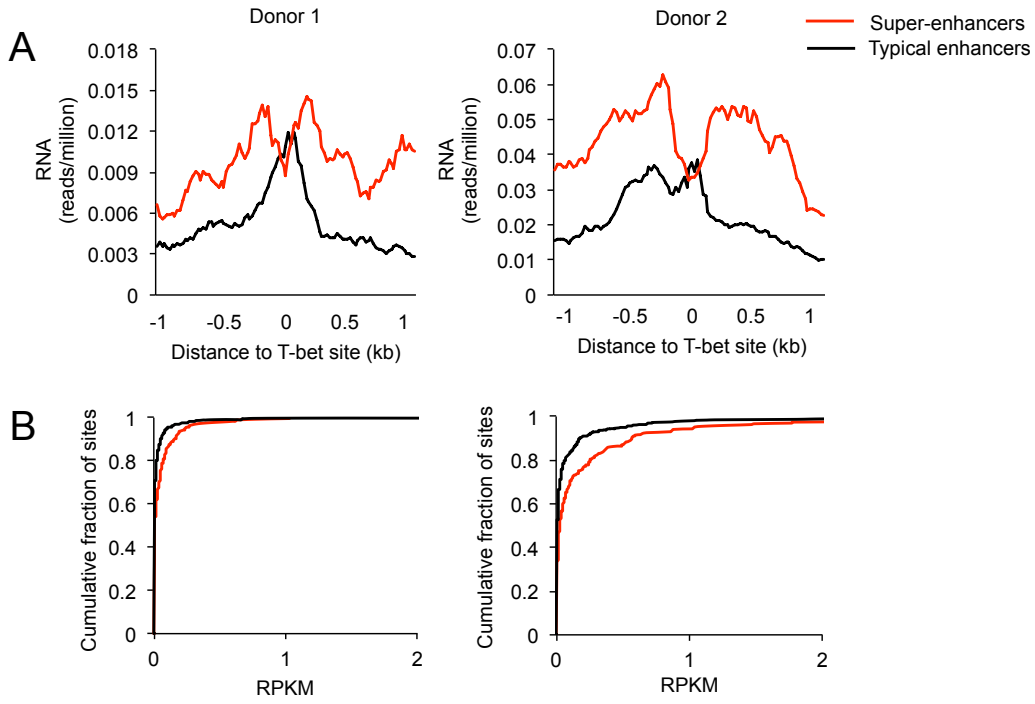
Figure 4.9: Super-enhancer eRNAs RNA are not polyadenylated

Comparison of RNA production in T-bet super-enhancers in Th1 cells from total and polyA+ purified fractions. A: Average profile of RNA (reads/million mapped reads) in super-enhancers centered on T-bet site locations. B: Cumulative frequency distributions for the amount of RNA produced from these sites. KS tests on distributions showed no statistical difference for donor 1 (D: 0.1041 P=0.102) while for donor 2 the maximum difference, D, was 0.1673 (P=0.001). RPKM: reads per kilobase per million mapped reads.

4.2.5 T-bet sites at super-enhancers produce more RNA than those at typical enhancers

As transcription at enhancers has been shown to be important for enhancer function [386, 421, 427] I sought to compare the amount of RNA produced from super-enhancers and typical enhancers. Super-enhancers are distinguished from typical enhancers by an increased presence of T-bet. This T-bet binding occurs over extended genomic regions with the boundary of the enhancer being defined by an absence of further T-bet binding in an additional 15kb window from the last peak. Super-enhancers are in essence more extended stretches of T-bet bound DNA and are therefore longer than normal enhancers. To account for this, rather than comparing the whole length of the enhancer for RNA production, I focussed on the region within close proximity of each T-bet binding site. To distinguish eRNAs from other transcripts, I excluded T-bet sites at enhancers that overlapped with known protein coding and lincRNA genes and repeats (tRNA, rRNA, snRNA). As Figure 4.10 shows, T-bet sites at super-enhancers were associated with increased production of RNA compared to those at typical enhancers. This is demonstrated by the average profile of RNA over the genomic site in Figure 4.10.A and also through the cumulative distribution of RNA from T-bet sites, Figure 4.10. B. Comparison of the cumulative distributions of super-enhancers and typical enhancers showed that the maximum difference in distribution, D, was 0.2132 ($P < 10^{-3}$) for donor 1 and 0.2122 for donor 2 and that this increase in RNA production is statistically significant ($P < 10^{-3}$ for both donors). Figure 4.10.C. demonstrates that for both donors, super-enhancers were more likely to produce RNA than typical enhancers (65% and 70% for

super-enhancers compared to 45% and 50% for typical enhancers) and had a higher median RPKM (0.01 and 0.027 for super-enhancers compared to 0 and 0.003 for typical enhancers). Therefore increased T-bet binding correlates with increased transcription.



C

	Donor 1		Donor 2	
	Tbet sites		T-bet sites at	
	Super-enhancers	Typical enhancers	Super-enhancers	Typical enhancers
Total number	269	908	269	908
Total number producing any RNA	176	408	187	454
Percentage producing any RNA	65.43	44.93	69.52	50.00
Median RPKM	0.00983	0.00000	0.02654	0.00283

Figure 4.10: Super-enhancers produce more RNA than typical enhancers
Comparison of amount of RNA produced from super-enhancers and typical enhancers. A: Average profile of RNA (reads/million mapped reads) at T-bet sites and super-enhancers compared to typical enhancers. Profiles are centered on T-bet sites and show a flank of 1kb each side. B: Cumulative frequency distributions for amount of RNA measured as RPKM produced from super-enhancers and typical enhancers. KS tests on distributions showed a maximum difference of 0.2132 ($P<10^{-3}$) and 0.2122 ($P<10^{-3}$) for donor 1 and 2 respectively. C: Comparison of T-bet sites at super-enhancers and typical enhancers in terms of the proportion that exhibit RNA production and their median RPKM. RPKM: reads per kilobase per million mapped reads.

4.2.6 Super-enhancer transcription is bidirectional

Several publications have reported bidirectional transcription from loci with potential roles in regulating gene expression. This includes short RNAs produced from transcription start sites [412, 413] and promoter upstream transcripts (PROMPTs) which are commonly found 0.5-2.5kb upstream of protein coding genes [417, 418, 434] and some descriptions of enhancer RNAs [420-423]. Bidirectional transcription has not however been previously examined at super-enhancers. The super-enhancer images presented in Figure 4.8 indicate that some transcript start sites exhibit this bidirectional character. Clear examples can be seen in the super-enhancers upstream of *GZMB* and *FURIN* (Figure 4.8). I therefore sought to quantify the bidirectional character of super-enhancer transcription using the same 1kb flanking regions around T-bet sites as described in Sections 4.2.2 to 4.2.5. Bidirectional transcription was mathematically defined as a lower than 2-fold difference in RNA-Seq read count between the regions up and downstream of a T-bet site (i.e. equal transcription each side of a T-bet site, log2 ratio of -1 to +1, Figure 4.11). Transcription with directionality was defined as a greater fold difference (log2 ratio of <-1 or >+1). As Figure 4.11 shows, I compared transcription from T-bet sites at super-enhancers with typical-enhancers and with the transcription start sites of RefSeq annotated genes, which are expected to show directional transcription. As expected, transcription from annotated genes exhibited clear directionality, shown by a high proportion of TSSs with an unequal up versus downstream read count ratio. Only 19% of TSSs showed unbiased directionality in their transcription while 81% showed a read count ratio of more than 2. Super-enhancers and typical enhancers differed from gene

TSSs with around a third of sites exhibiting equal directionality in transcription. This is a similar proportion of bidirectional transcription as published for other enhancers [421]. Super-enhancers and typical enhancers showed comparable proportions of bidirectional transcription (35% and 33%, respectively).

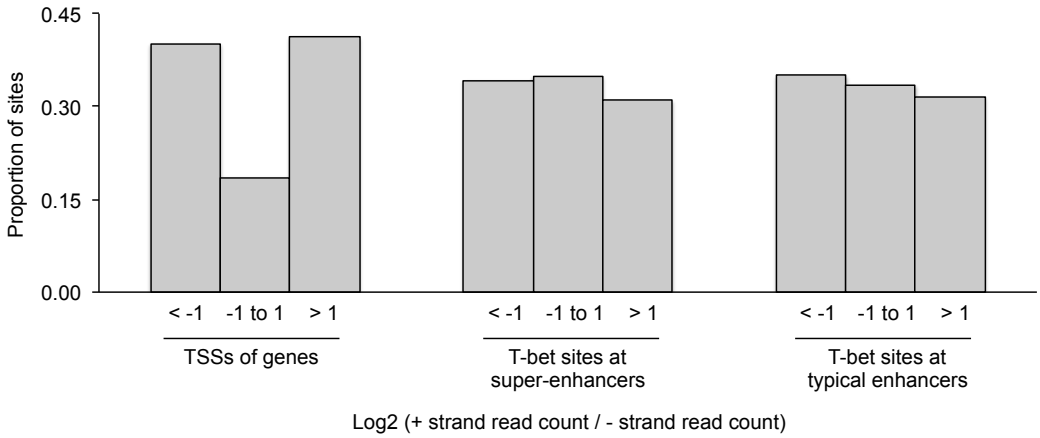


Figure 4.11: Transcription from T-bet sites at super-enhancers is bidirectional
The directionality of transcription around T-bet sites in super-enhancers and typical enhancers was quantitated by calculating the ratio of reads upstream (+ strand) and downstream (- strand) of each site. Bidirectionality was defined as a fold difference of less than 2 (log2 ratio between -1 and +1). The proportion of T-bet sites associated with bidirectional transcription and with directional preference was calculated. The start sites of RefSeq annotated genes (genes) are given as a comparator expected to have directional preference. TSSs: transcription start sites.

4.2.7 T-bet is required for IFNG super-enhancer transcription

Our lab and others have described elements around the *Ifng* locus that are highly conserved and bound by T-bet [245, 246, 337-340]. These sites are DNaseI hypersensitive and marked by H3K4me1, which are features of enhancers [246, 344, 346]. Similar sites are seen genome-wide at locations distal to Th1-specific genes with T-bet binding being required for expression of these genes [246]. Figure 4.12 illustrates the position of T-bet sites distal to the *Ifng/IFNG* locus in mouse and human showing their locations relative to the IFNG super-enhancer and RNA sequencing data in human. In other literature these conserved non-coding sequences have been referred to by their locations in mouse relative to Ifng: CNS-6 (Human -3.8kb in Figure 4.12), CNS-22 (-15.7kb and -16kb), CNS-34 (-30kb and -30.4kb).

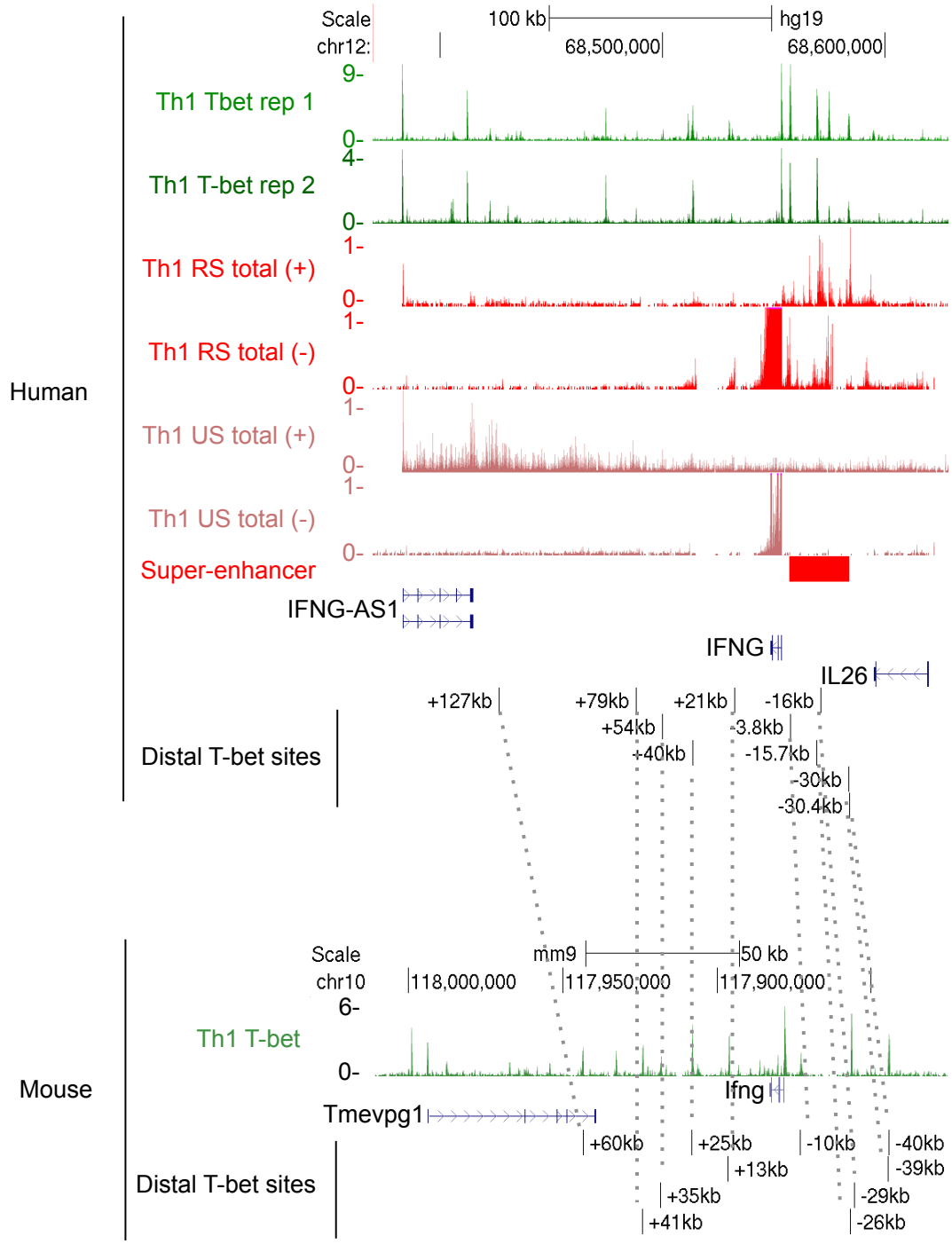


Figure 4.12: T-bet sites distal to Ifng/IFNG in mouse and human

T-bet binding sites distal to Th1-specific genes have been previously published and are exemplified by those at the Ifng/IFNG locus. These binding sites are conserved between mouse and human and are illustrated here in relation to the IFNG super-enhancer and RNA sequencing data. Dotted lines link up corresponding conserved sites. The murine locus is shown in its reverse orientation. +: positive strand; -: negative strand.

To test if this T-bet binding was necessary for eRNA transcription from the *IFNG* super-enhancer, naïve CD4 T cells from wild type and T-bet knockout ($Tbx21^{-/-}$) mice were cultured under Th1 and Th2 polarising conditions and the level of eRNA at these sites assess by qPCR (Figure 4.13, cultures performed by Ian Jackson). Similar to the trend visible through RNA sequencing (Figure 4.8), eRNAs transcribed from the super-enhancer (-40kb, -38.7kb, -29kb, -26kb and -10kb) were increased in Th1 compared to Th2 cells and were transcribed at a greater level in restimulated compared to unstimulated cells. This pattern was reflected in the expression of *Ifng* itself. In contrast, most downstream transcripts were decreased with restimulation and naïve Th1 progenitor cells showed a higher level of -26kb transcript than Th1 cells (Figure 4.13), demonstrating the complexity of non-coding transcription that may be involved in regulating Ifng expression. With loss of T-bet, expression of all these targets was reduced. This was also true of transcripts downstream of *Ifng* (+25kb, +35kb, +41kb, +60kb) while the housekeeping gene Gapdh did not show this trend Figure 4.13). This demonstrated that full transcription of eRNAs from the *IFNG* super-enhancer, and other enhancers at the *Ifng* locus, requires T-bet binding.

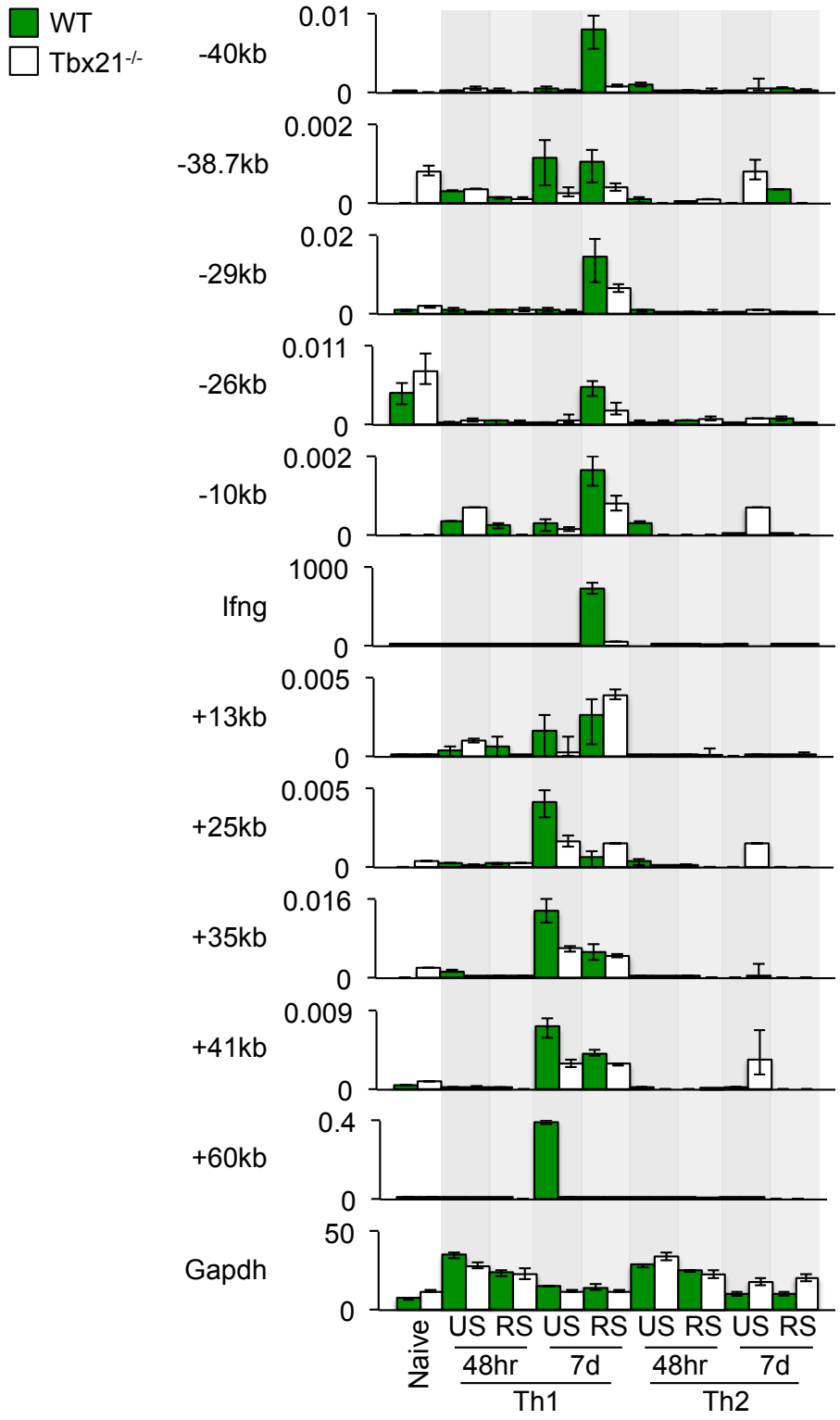


Figure 4.13: T-bet is required for Ifng super-enhancer transcription

Naïve T cells from WT and T-bet knockout (Tbx21$^{-/-}$) mice were cultured under Th1 and Th2 polarising conditions and harvested at 48 hours and 7 days with and without restimulation. QPCR was performed for Ifng and several eRNA loci up and downstream. Targets are shown in order of their distance to Ifng and named with reference to this. Ifng and eRNAs upstream showed an increase upon stimulation, which was limited upon loss of T-bet. Downstream eRNAs however, while their full expression requires T-bet, are decreased upon stimulation. The housekeeping gene *Gapdh* does not show these trends. Enrichment is normalised to Hprt**.** Data are mean and standard deviation. US: unstimulated; RS: restimulated.

4.2.8 IFNG upstream eRNA transcription is dependent upon P-TEFb

As described, our lab has observed that transcriptional regulation of lineage-specific gene expression occurs at the point of elongation rather than initiation (Figure 4.2.A, Richard Jenner). ChIP-Seq of the elongation factor P-TEFb shows that lineage-specific processivity of RNAPII and recruitment of P-TEFb is enriched at genes associated with super-enhancers compared to those at typical enhancers, implying that the greater T-bet binding seen at super-enhancers is associated with increased transcriptional elongation through P-TEFb recruitment (Figure 4.2.B, Richard Jenner). Supporting this, the coincident localisation of T-bet and P-TEFb at Th1 loci can be seen in Figure 4.8. As greater RNA production is also seen at super-enhancers compared to typical enhancers (Figure 4.10), I hypothesised that P-TEFb binding at super-enhancers is required for super-enhancer eRNA transcription. To test this, mouse naïve T cells were cultured under Th1 and Th2 polarising conditions with and without treatment with inhibitors of P-TEFb (Arnulf Hertweck). Two P-TEFb inhibitors were used: JQ1, which blocks the histone acetyl recognising BET domain of BRD4 that functions in P-TEFb recruitment [381]; and Flavopiridol, which inhibits the kinase activity of the Cdk9 subunit of P-TEFb that phosphorylates RNAPII [380].

Figure 4.14 displays the effect of inhibition of P-TEFb recruitment with JQ1 while Figure 4.15 shows the effect of inhibition of P-TEFb activity with Flavopiridol. As expected *Ifng* expression and transcripts around the Ifng locus were increased in Th1 compared to Th2 cells. Consistent with experiments with human cells (Figure 4.7, Figure 4.12), restimulation of

mouse Th1 cells also led to an increase in eRNA upstream of *Ifng* while transcripts further downstream (+60kb in particular) were decreased upon restimulation (Figure 4.12). The increase in transcription of *Ifng* and eRNAs is reduced in the presence of both JQ1 (Figure 4.14) and Flavopiridol (Figure 4.15) with dosage dependent effect observed. Restimulation and treatment with inhibitors did not affect the expression of the housekeeping gene *Gapdh* the expression of which is not lineage specific. Therefore the T-bet associated upregulation of eRNAs at the super-enhancer upstream of *Ifng* requires P-TEFb function.

In addition, JQ1 and Flavopiridol differed in their effect on transcript production downstream. While JQ1 had the same dose dependent effect as at upstream targets and high concentration of Flavopiridol (10μM, Figure 4.15 [++]) also inhibited transcription, treatment with a lower concentration of Flavopiridol (1μM, Figure 4.15 [+]) lead to a seemingly paradoxical increase in transcription. This occurred with restimulated but not unstimulated cells and only in Th1-polarised cells. This reinforces the concept that regulation of non-coding transcripts around the Ifng locus is complex with transcripts upstream and downstream showing differential response.

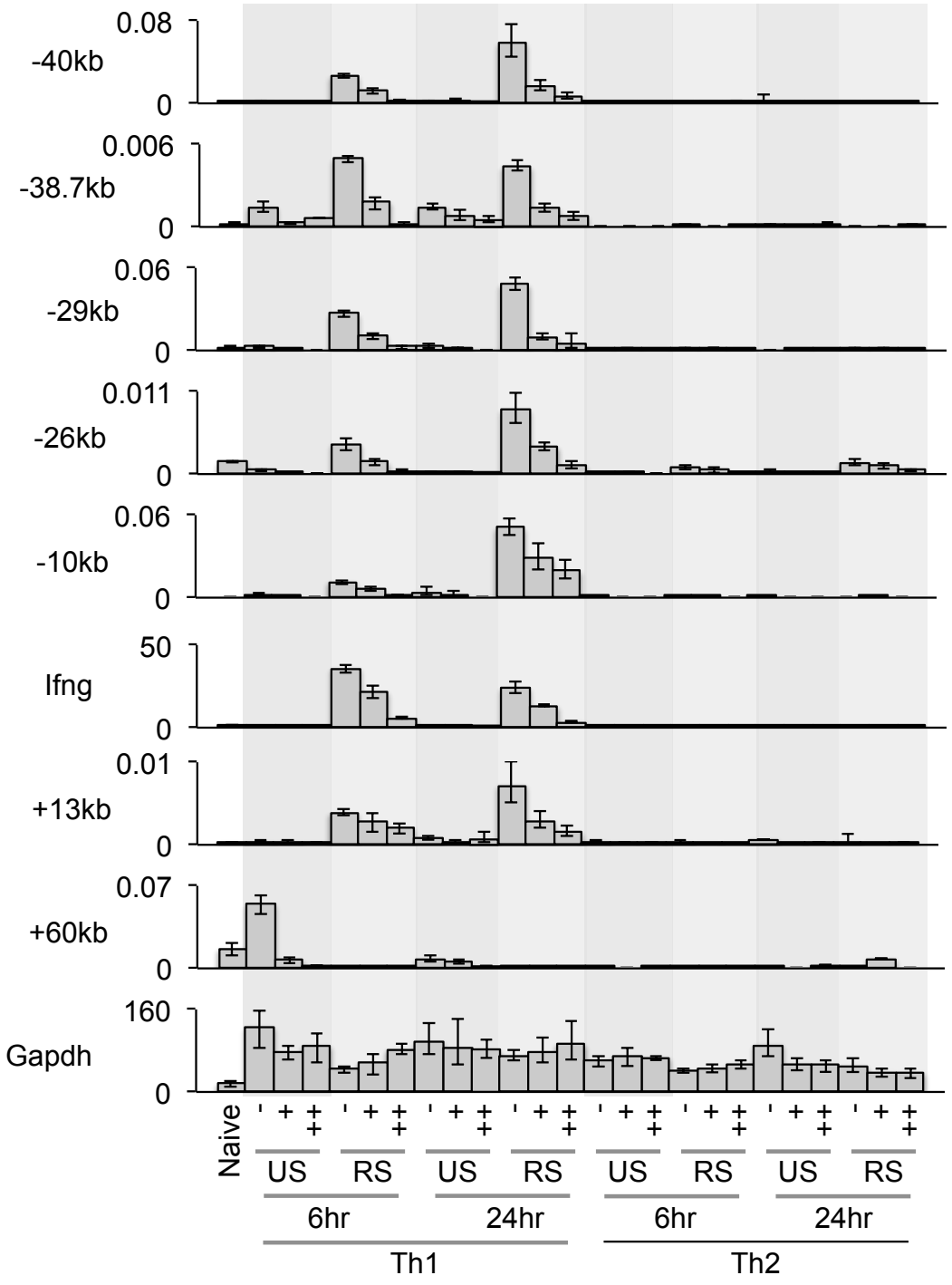


Figure 4.14: P-TEFb inhibition with JQ1 diminishes eRNA transcription

Naïve T cells were cultured *in vitro* under Th1 and Th2 polarising conditions for 13 days before treatment with the P-TEFb inhibitor JQ1. P-TEFb inhibition was performed with and without restimulation, for either 6 or 24 hours, with low or high concentrations of JQ1 or none at all. QPCR was performed for Ifng and several eRNA loci up and downstream. Targets are shown in order of their distance to Ifng and named with reference to this. Ifng and eRNAs upstream showed an increase upon stimulation, which was limited with addition of JQ1 in a dose dependent manner. Due to limited material, two downstream targets were prioritised for analysis and showed the expected expression given the trend observed in Figure 4.13. The housekeeping gene Gapdh does not show these trends. Enrichment is normalised to Hprt**.** Data are mean and standard deviation. US: unstimulated; RS: restimulatedJQ1. -: DMSO only; +: 50nM JQ1; ++: 500nM JQ1.

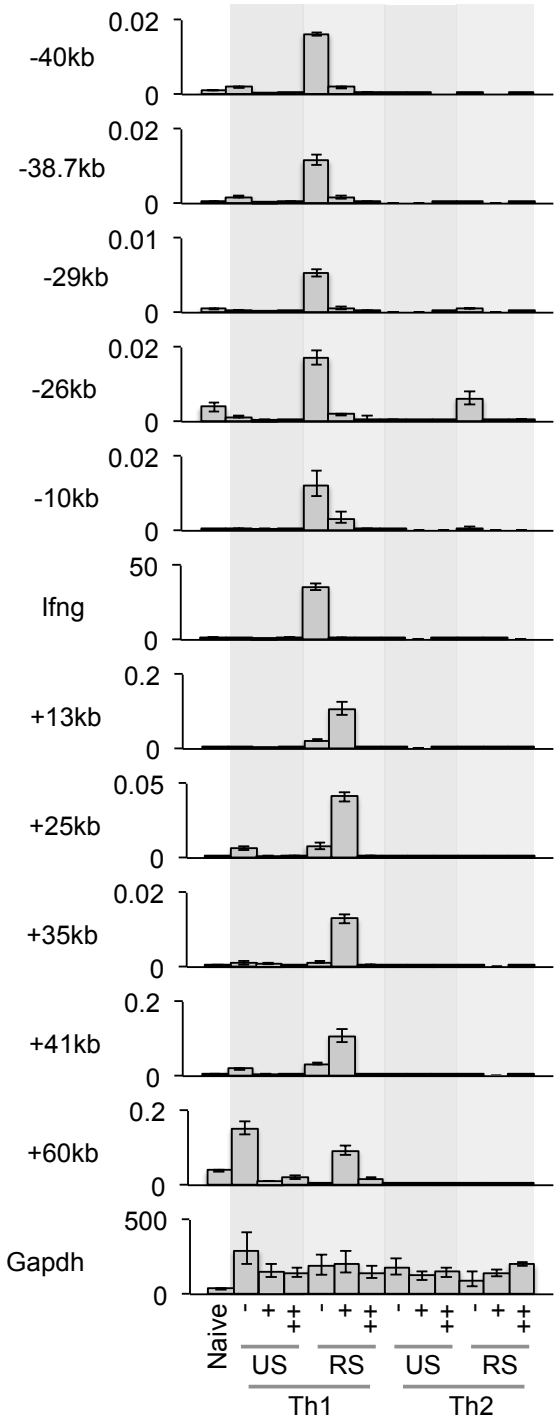


Figure 4.15: P-TEFb inhibition with Flavopiridol diminishes eRNA transcription
Naïve T cells were cultured *in vitro* under Th1 and Th2 polarising conditions for 13 days before treatment with the P-TEFb inhibitor Flavopiridol. P-TEFb inhibition was performed with and without restimulation with low or high concentrations of Flavopiridol or none at all. QPCR was performed for Ifng and several eRNA loci up and downstream. Targets are shown in order of their distance to Ifng and named with reference to this. Ifng and eRNAs upstream showed an increase upon stimulation, which was limited with addition of Flavopiridol in a dose dependent manner. Downstream targets showed the same trend observed in Figure 4.13 regarding restimulation and exhibited an unexpected increase with the lower concentration of Flavopiridol. The housekeeping gene Gapdh does not show these trends. Enrichment is normalised to Hprt**.** Data are mean and standard deviation. US: unstimulated; RS: restimulated. -: DMSO only; +:1μM Flavopiridol; ++:10μM Flavopiridol.

4.2.9 Characterisation of novel T-bet super-enhancer eRNAs

In order to identify novel specific transcripts, RNA-Seq data was assembled into transcript structures using Cufflinks (John Ambrose) and transcripts at T-bet super-enhancers identified as described in Section 2.5.9. RNAs overlapping repeats (tRNA, rRNA and snRNA) and the coding regions of Gencode annotated transcripts were excluded. As eRNAs do not exhibit splicing, multi-exonic transcripts were also excluded. The identities of all eRNAs at T-bet super-enhancers are described in Table 7.30. This table details the coordinates and strand of each transcript and the super-enhancer it is transcribed from, which have been designated IDs based on their nearest protein-coding gene. The level of each transcript (FPKM) in each condition for donor 2 is also given. As expected, transcripts at T-bet super-enhancers are transcribed at a higher level in total RNA from restimulated Th1 cells compared to polyadenylated fractions, Th2 cells and unstimulated cells, showing that the general trends observed genome wide in Figure 4.6, Figure 4.7 and Figure 4.9 are features of each individual transcript. Twenty-one transcripts were identified at the *IFNG* super-enhancer. These are described in Table 7.30 and illustrated in Figure 4.16.

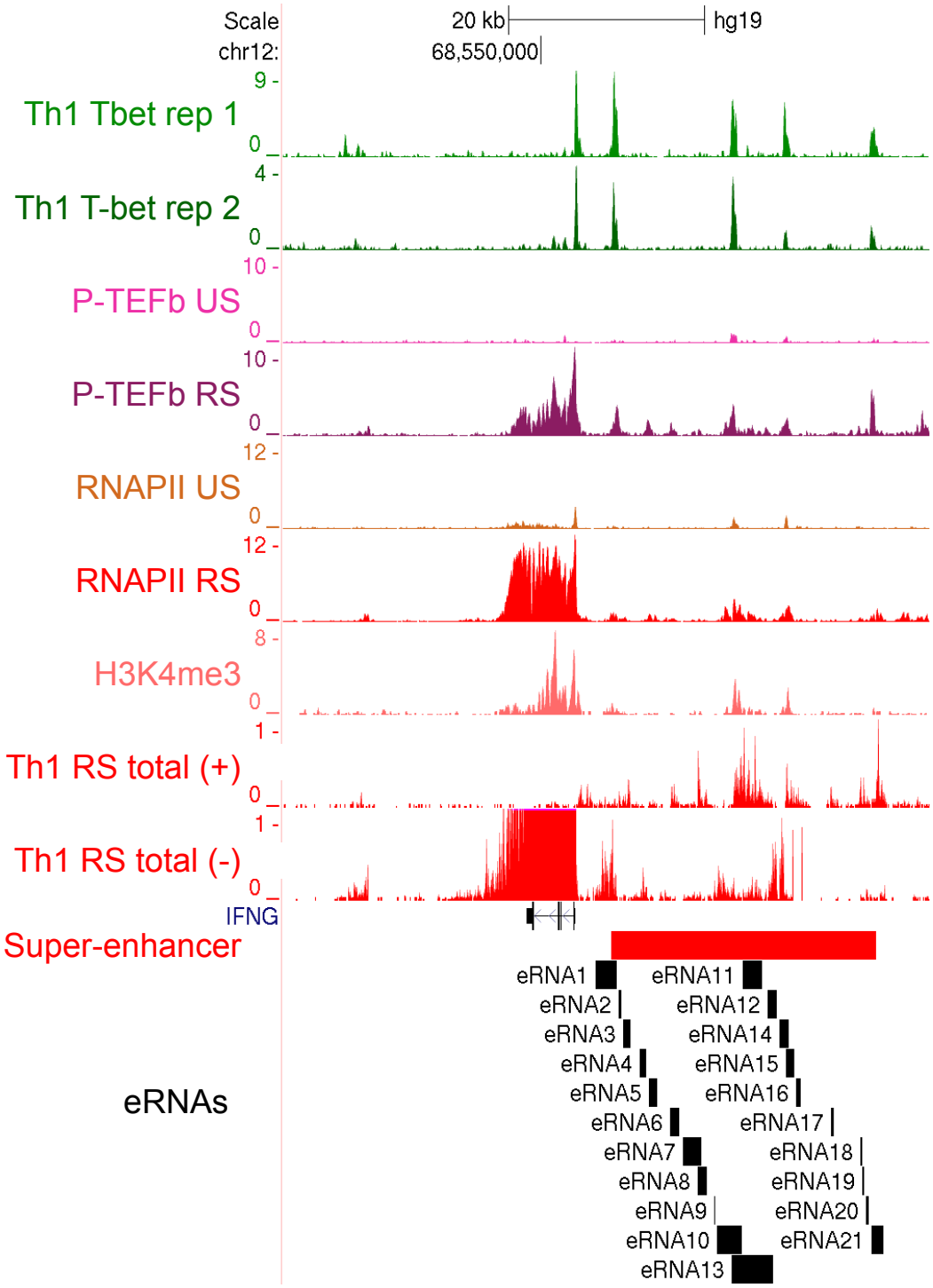


Figure 4.16: Novel eRNAs at the IFNG super-enhancer

Locations of individual eRNA transcripts at the IFNG super-enhancer as detailed in Table 7.30. Their positions are shown relative to T-bet, P-TEFb and RNAPII binding sites and trimethylated H3K4. +: positive strand; -:negative strand. RS: restimulated; US: unstimulated.

4.3 Discussion

Observations in our lab have has shown that lineage-specific gene expression during Th1 versus Th2 differentiation is primarily regulated at the point of transcriptional elongation rather than initiation (Figure 4.2.A). In addition, lineage-specific genes that exhibit increased transcriptional elongation are preferentially associated with super-enhancers (Figure 4.2), which are defined by increased binding of the lineage-specific transcription factor T-bet compared to typical enhancers.

As the function of enhancers has been associated with transcription of ncRNAs, termed eRNAs [421, 425, 428, 431], I investigated the transcriptional output of super-enhancers. I report here that in addition to super-enhancer associated genes the super-enhancers themselves exhibit lineage-specific transcription (Figure 4.6) and transcription is increased upon restimulation (Figure 4.7), showing that transcription from super-enhancers mirrors that of associated protein coding genes. In addition, this super-enhancer transcription shows characteristics of eRNA that have been associated with functional effects, including lack of polyadenylation (Figure 4.9, [420, 421]) and bidirectional transcription (Figure 4.11 [420, 421]). Therefore the T-bet super-enhancers eRNAs identified here appear to be regulated and processed in the same manner as previously identified functional eRNAs. I have identified individual eRNAs produced from the T-bet super-enhancer upstream of *Ifng* (Table 7.30) and at other T-bet super-enhancers throughout the human genome (Table 7.30). As expected these are higher in Th1 restimulated cells compared to Th2 cells and unstimulated cells (Table 7.30). Furthermore, super-enhancers produce

more RNA than typical enhancers (Figure 4.10), which is consistent with the production of ncRNA being a functional effect of the increased T-bet occupancy that defines super-enhancers. Supporting this, loss of T-bet leads to reduced transcription of eRNAs from the Ifng super-enhancer and of other potential regulatory transcripts downstream of the *Ifng* locus (Figure 4.13), suggesting that this transcription is a direct result of T-bet binding. The ultimate proof of a functional effect for these eRNAs would be to knock them down or delete them without affecting DNA-dependent enhancer effects and demonstrating they are necessary for full expression of their associated genes.

Lineage-specific gene expression is controlled at the point of transcriptional elongation and at super-enhancers T-bet binding is colocalised with the elongation factor P-TEFb (Figure 4.8). I hypothesised that the elongation factor P-TEFb may have functional effects at super-enhancers through transcription of eRNAs. Treatment with two inhibitors of P-TEFb, which block activity at different points, reduced transcription of eRNAs at the *Ifng* super-enhancer (Section 4.2.8, Figure 4.14 and Figure 4.15) suggesting that P-TEFb binding at T-bet sites is functional. This is consistent with lineage-specific control of gene expression by super-enhancers being mediated through recruitment of P-TEFb by T-bet.

The presence of H3K4me3 at enhancers has been debated [293, 344]. However the super-enhancer eRNA loci I report here are associated with H3K4me3 (Figure 4.8). eRNA loci have instead been reported to exhibit higher levels of H3K4me1 than H3K4me3 [293, 344, 356]. As I lack H3K4me1 data for human Th1 cells I cannot confirm this. However it is

interesting to note that at the *IFNG* locus the level of H3K4me3 is comparable between the gene itself and the eRNAs upstream (Figure 4.8). At other super-enhancers presented here (Figure 4.8) there is greater enrichment of H3K4me3 at the gene compared to eRNAs. This might suggest differences in regulation or activity of eRNAs at different genes.

Non-coding transcripts downstream of Ifng (+25kb, +35kb, +41kb and +60kb,Figure 4.13,Figure 4.14 and Figure 4.15) were regulated differently to those at the Ifng super-enhancer (-40kb, -29kb, -25kb, -10kb,Figure 4.13,Figure 4.14 and Figure 4.15) with restimulation leading to a decrease in the level of RNA as assessed by qPCR (Figure 4.13). This response to stimulation in this region can also be seen through RNA sequencing (Figure 4.12). Corroborating this, Tmevpg1, a long ncRNA, which promotes Ifng expression in mice, has been previously observed to be downregulated in response to stimulation [463]. This may reflect differences in transcriptional regulation during polarisation compared to those that occur upon restimulation. The exception to this trend is the +13kb transcript, which is downstream of *Ifng* but is also upregulated upon stimulation (Figure 4.13, Figure 4.14 and Figure 4.15). As Figure 4.12 shows, this transcript corresponds to +21kb location in humans, which is bound by T-bet in our ChIP-Seq data. In Figure 4.8 P-TEFb can be seen to colocalise with T-bet. This shows that at this location downstream of *Ifng* there is the same relationship between T-bet, P-TEFb and production of RNA as at the *Ifng* super-enhancer. This common regulation may result from the *Ifng* super-enhancer and the downstream region immediately flanking the Ifng gene being in close physical proximity, which is suggestive of DNA looping. Binding of CTCF, a chromatin insulator

involved in DNA looping, has been previously described around the *IFNG* locus at several sites [230]. An interaction between the published human -31kb and +22kb CTCF locations **[230]** would explain the common regulatory events stretching from the *IFNG* super-enhancer (starting around -30kb in human) to the immediate downstream site (+21kb in human). The insulating effect of CTCF explains why the same trend in regulation with stimulation does not extend beyond this region downstream of Ifng and into the IFNG-AS1 (Tmevpg1) locus.

A difference was also seen between eRNA regulation and all downstream transcripts in a paradoxical increase in transcription downstream of Ifng upon treatment of Flavopiridol upon stimulation (+13kb, +25kb, +35kb, +41kb, +60kb, Figure 4.15). This could be explained by an accumulation of these pre-transcribed RNA at a post-transcriptional level or off-target effects of Flavopiridol that increase transcription. This upregulation in transcript level at particular sites with Flavopiridol treatment has been reported in human fibroblasts [563] with increased RNAPII occupation at gene bodies also noted showing that this effect is due to increased transcriptional elongation. This suggests a P-TEFb independent mechanism for elongation of RNAPII at some loci, which is supported by an observation that inhibition of Cdk9 with another inhibitor, DRB, has limited effect on the level of some transcripts [386]. In addition, as JQ1 did not have the same effect as Flavopiridol, this effect may be a response to Cdk9 inhibition, which is targeted by Flavopiridol but not JQ1. In addition to its role in phosphorylating RNAPII, Cdk9 has addition roles including in guiding chromatin modifications [564]. In this context of human fibroblasts, upregulation with Flavopiridol treatment was also seen at specific immune

responsive loci and was suggested to be a stress induced response to P-TEFb inhibition [563]. This could suggest the region downstream of Ifng responds differently to inflammatory signals than the upstream region allowing variable response to the locus depending on environmental stimulus.

The observed differences in regulation between super-enhancer eRNAs and those downstream of Ifng, including Tmevpg1, illustrate the complexity of ncRNA transcription around the Ifng locus.

5 Functional roles of the identified lineage specific of ncRNAs

5.1 Introduction

I have identified ncRNAs that differ in expression between T cell lineages and may have functional implications for lineage specification. This includes the identification of long intergenic ncRNAs preferentially expressed in primary human Treg compared to Tresp (Chapter 3) and enhancer RNAs with Th1 specific expression (Chapter 4). A key question is whether or not these ncRNAs are functional, whether instead it is the act of transcription of the RNAs that is important rather than the RNA itself, or whether the RNAs are merely transcriptional “by-products” with no role in the cell.

Examination of my Treg lincRNA data and the surrounding literature suggested two RNAs were the best candidates for having a function in Treg biology. CRNDE is upregulated in colorectal cancer [524] and other cancers [524, 548] and the high expression in ESCs [442] and the decrease in expression with specialisation [526] have led to an association of CRNDE with a cycling stem-cell state. An unspliced *CRNDE* transcript containing a highly-conserved intronic region has been shown to promote growth and suppresses apoptosis in fibroblasts [526]. Additional unspliced transcripts are downregulated by insulin signalling through mTOR and Raf/MAPK [550] and knockdown of these transcripts impacts upon processes downstream of insulin such as glucose and lipid metabolism [550]. A physical association has been identified between CRNDE and the chromatin-modifying complexes PRC2 and CoREST using native RNA IP

[441], leading to the suggestion that the CRNDE transcript may target recruitment of these complexes with effects on regulation of transcription.

A second candidate is LOC286442. While there is no published data on the function of this uncharacterised lincRNA, LOC286442 had the highest fold-difference between Treg and Tresp of any identified lincRNA by array (13.9 fold, Figure 3.3). In addition, FOXP3 binding was apparent at the promoter and in intronic regions (Figure 3.6), supporting Treg-specific regulation of the *LOC286442* gene.

Enhancer RNAs have been found to contribute to enhancer function in other cell types. Transcription from T-bet super-enhancers is increased compared to typical enhancers (Figure 4.10) and this transcription is Th1 specific (Figure 4.6). RNAPII binding co-localises with the Th1 lineage-specifying transcription factor T-bet and the transcriptional elongation factor P-TEFb, with loss of T-bet and inhibition of P-TEFb limiting super-enhancer eRNA transcription (Figure 4.13 and Figure 4.14, respectively). This suggests T-bet may function at super-enhancers to recruit P-TEFb to lineage specific genes. I have identified the individual transcripts produced from the super-enhancer upstream of *IFNG* and hypothesised that super-enhancer eRNAs also have a functional impact on gene expression.

5.1.1 Knockdown strategies to interrogate ncRNA function

One method to interrogate ncRNA function is through knockdown approaches, using RNA interference (RNAi). Figure 5.1 illustrates the mechanisms and cellular machinery involved in knockdown of coding and non-coding transcripts through RNAi in order to assess how their loss

affects cellular function. These mechanisms include short interfering RNA (siRNA) and short hairpin RNA (shRNA), which require delivery into the cell through transfection or vector (including viral vectors) transduction, respectively. Both mechanisms rely upon the cellular machinery that processes and enacts endogenous post-transcriptional gene silencing (PTGS) through genomically encoded microRNAs (miRNA). The requirement for cytoplasmic processing machinery potentially limits the effectiveness of shRNA for knockdown of non-coding transcripts that may be localised to the nucleus. As shown in Figure 5.1, cell intrinsic RNA interference is mediated by genomically encoded microRNAs (miRNA). miRNAs are transcribed as long primary transcripts called pri-miRNAs that are then processed by Drosha in the nucleus [565, 566] to the pre-miRNA form, which consists of a stem-loop structure formed by complementary base pairing. This pre-miRNA is a substrate for export to the cytoplasm by Exportin5 [567], where it is further processed by Dicer [568] to a mature miRNA for loading into the Argonaut/RISC (RNA induced silencing complex) complex [569, 570]. Argo/RISC catalyses the separation of the passenger (sense) guide (anti-sense) miRNA strands and guides the annealing of the passenger strand with its complementary seed region in the 3' untranslated region (UTR) of the target mRNA. With full complementarity this annealing leads to transcript accumulation within P-bodies where degradation occurs [571, 572], while partial complementarity leads to translational repression [571], both of which ultimately lead to reduced expression of the protein encoded by the target mRNA. Figure 5.1 illustrates how shRNA, which resembles the pri-miRNA form, co-opts this process. siRNAs can be introduced into cells through transfection, either

using electroporation/nucleofection or using liposomes, and already being in a mature, pre-processed form, siRNAs are directly loaded into Argonaut/RISC complex in either the cytoplasm or nucleus [573]. A potential advantage of using siRNA is that it by-passes the requirement for Dicer, which is generally thought to be restricted to the cytoplasm [574].

5.1.1 Locked nucleic acids as an alternative strategy to knockdown ncRNAs

Several publications have recently described the use of locked nucleic acids (LNA) for the knockdown of ncRNA targets ([421, 427]). LNAs contain a ribose moiety with a covalent bond between the 2' and 4' oxygens [575, 576]. This bridge locks the ribose conformation, enhancing the stability, specificity and sensitivity of base pairing with a target [577]. LNAs can also contain DNA or RNA nucleotides, as either mixmers, in which the LNAs are interspersed with the regular nucleotides, or gapmers, in which the LNAs flank a stretch of DNA or RNA. Annealing of an antisense LNA-DNA gapmer with a target complementary RNA can be used to degrade target RNAs through enzymatic cleavage by RNaseH, which recognises DNA:RNA hybrids [578]. In addition, LNA hybridisation may sterically block the activity of a transcript [579]. Due to the stability of LNA, entry into a cell can be mediated through gymnosis [580] i.e. entry into the cell in parallel with nutrients during normal cell growth.

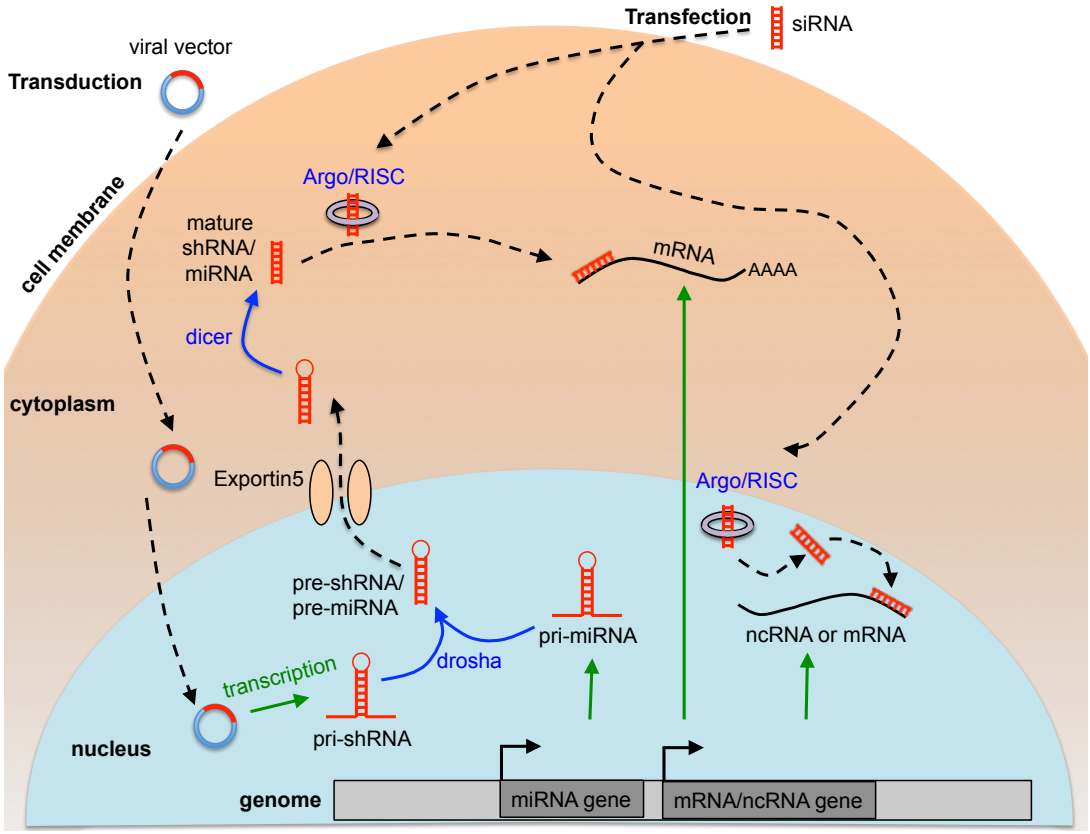


Figure 5.1: Knockdown mechanisms mediated by cellular RNAi machinery
RNA interference mechanisms include endogenous microRNA (miRNA), short hairpin RNA (shRNA) and short interfering RNA (siRNA). Pri-pre-miRNA are processed by Drosha in the nucleus to their pre-miRNA form, which is exported to the cytoplasm by Exportin5. Here, further processing is performed by Dicer to form a mature miRNA for loading into the Argonaut/RISC complex. Argo/RISC catalyses the separation of the passenger (sense) guide (anti-sense) miRNA strands and guides the annealing of the passenger strand with its complementary seed region in the 3' untranslated region (UTR) of the target mRNA. shRNA, which is transcribed in the nucleus after transduction of a viral vector co-opts this process. Short interfering RNA however, enters the cell through electrical or chemical induction of membrane pores through transfection, and being in a mature, pre-processed form, is directly loaded into Argonaut/RISC complex in either the cytoplasm or nucleus. RISC: RNA induced silencing complex.

5.1.2 Aims

In the experiments below I have attempted to explore the functional roles of the Treg lincRNAs I identified in Chapter 3 and of the super-enhancer eRNAs I identified in Chapter 4. I first sought insights into the biological roles of the Treg lincRNAs though gene ontology of co-regulated transcripts identified by hierarchical clustering. I then optimised knockdown approaches in order to investigate the effect of the loss of these lincRNAs on Treg phenotype and super-enhancer eRNAs on downstream gene expression using the IFNG locus as a model.

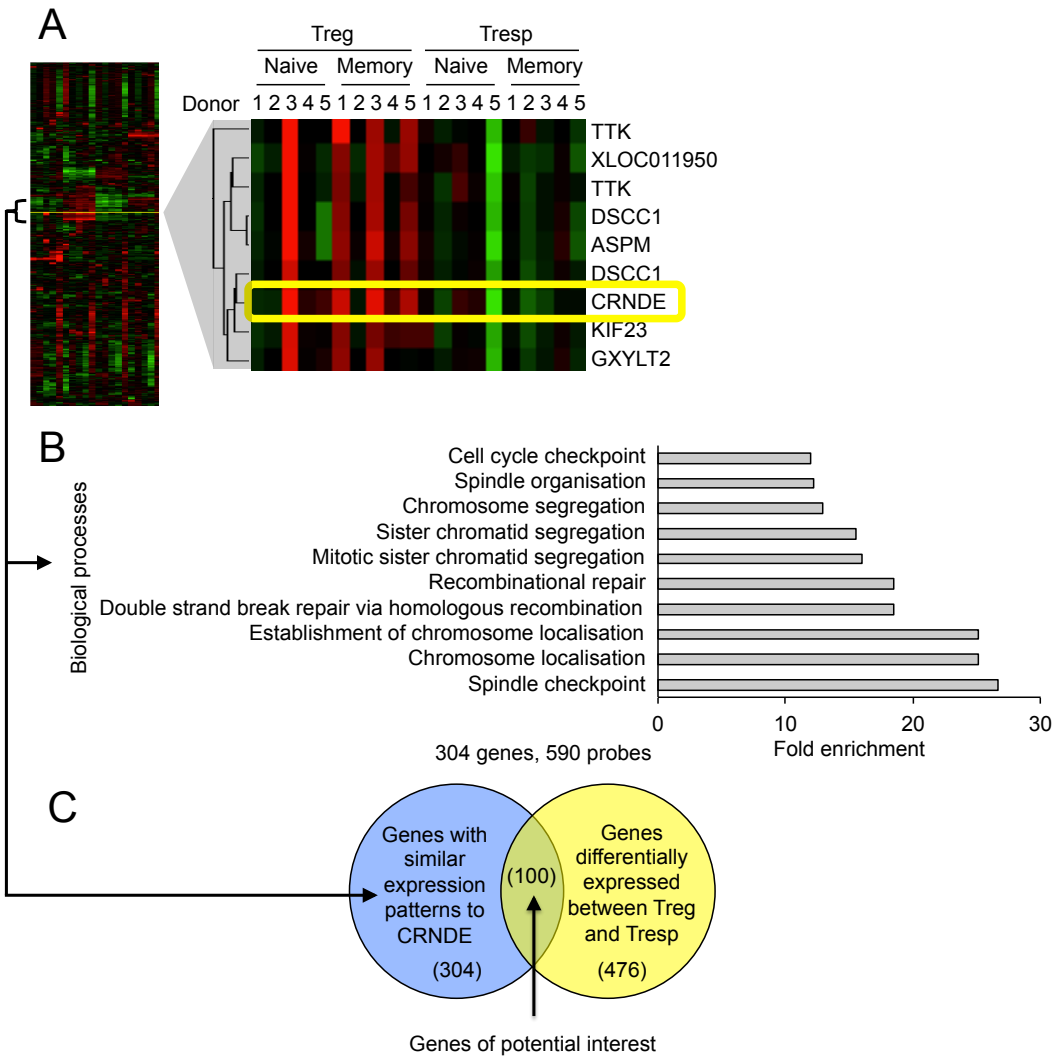
1. To use Gene Ontology to identify potential biological roles for Treg lincRNAs based on the co-regulated genes with which they cluster.
2. Optimise strategies to knock-down lincRNAs in Tregs using siRNAs.
3. Optimise the use of LNA oligonucleotides for interrogating eRNA function in Th1 cells.

5.2 Results

5.2.1 Genes and biological processes associated with CRNDE

In order to identify biological process that are relevant to CRNDE function, I examined genes with known functions that exhibited expression patterns similar to *CRNDE*. This "guilt-by-association" approach has previously been used to provide insights into potential functions for unannotated genes [441]. Genes were identified by hierarchical clustering of microarray data (Section 2.5.2). As Figure 5.2.A illustrates, the genes with the most similar expression, as shown by their proximity in the cluster, included *TKK* (TTK protein kinase [MSP1]), a spindle checkpoint kinase [581], *DSCC1* (Defective in sister chromatid cohesion 1), an S phase replication factor [582], *ASPM* (abnormal spindle homolog), *KIF23* (kinase family member 23) and *GXYLT2* (glucoside xylosytransferase 2). This small cluster therefore appeared relevant to cell cycling. A larger cluster of genes allowed identification of statistically significantly associated biological processes through gene ontology (304 genes, P (adjusted) <0.05). These biological processes also included cell cycling thereby also reflecting the local cluster of genes Figure 5.2.B. To select for genes with the most relevance to Treg biology, I identified genes in this cluster that were also differentially expressed between Treg and Tresp (Figure 3.2.A, Table 7.1 and Table 7.2). From this list of genes (Table 7.31) those of particular relevance to Treg biology included *TNFRSF8* (also known as CD30) and *TNFRSF9* (known as CD137 and 4-1BB). CD30 is preferentially expressed on activated compared to resting lymphocytes, in agreement with the array data (Figure 5.2) and is relevant to Treg function in the context of

autoimmunity through limitation of autoreactive CD8 T cell proliferation [583], while signalling via 4-1BB promotes Treg expansion [221, 584, 585]. MKI67 (ki67) was also of interest because, although its expression is not specific to Treg function, it is commonly used as a proliferation associated antigen in Treg assays. I hypothesised that CRNDE may have a role is these processes through regulating the expression, directly or indirectly of ki67, CD30 and/or 4-1BB.



Gene	*Relevance*	*References*
MKI67 (Ki67)	Upregulated during cell cycling	[586]
TNFRSF8 (CD30)	Expressed on activated lymphocytes. Has a role in autoimmunity through limiting autoreactive CD8 T cell proliferation	[583]
TNFRSF9 (4-1BB, CD137)	Downstream signalling leads to Treg expansion	[584]

Figure 5.2: Genes and biological processes associated with CRNDE expression. Genes with similar expression patterns to *CRNDE* as identified by hierarchical clustering were used in a 'guilt-by-association' approach to identify biological processes other genes with possible relevance to CRNDE function. A: The genes with most similar expression to *CRNDE*. B: A larger cluster of genes were used in gene ontology to identify biological processes associated with *CRNDE* expression (304 genes, P (adjusted) <0.05). C: These genes were compared with those differentially expressed between Treg and Tresp (476 genes, Figure 3.2.A, Table 7.1Table 7.2) to identify those with those most relevant to Treg biology. From this list (Table 7.31) MKI67, TNFRSF8 and TNFRSF9 were recognised as particularly interesting. TKK: TTK protein kinase; DSCC1: Defective in sister chromatid cohesion 1; ASPM: abnormal spindle homolog; KIF23: kinase family member 23; GXYLT2: glucoside xylosytransferase 2.

5.2.2 CRNDE expression in response to T cell activation

To better understand the association between *CRNDE* expression and T cell activation and proliferation, I examined its expression during *in vitro* T cell polarisation cultures using the microarray and RNA-Seq data obtained in Chapter 4. Figure 5.3 shows the differential expression between naïve T cells and Th1 and Th2 cells. *CRNDE* had a notably higher expression in *in vitro* polarised cells compared to unactivated *ex vivo* naïve T cells. This is shown in relation to changes in expression of other Treg ncRNAs (Figure 5.3.A). Most of these Treg ncRNAs did not show this trend. I confirmed this differential expression by qPCR (Figure 5.3.B) and RNA-Seq data (Figure 5.3.C). This supports a role for CRNDE in general T cell activation rather than a specific involvement in Treg phenotype. Additionally, CRNDE also appeared to be upregulated in Th2 cells compared to Th1 cells.

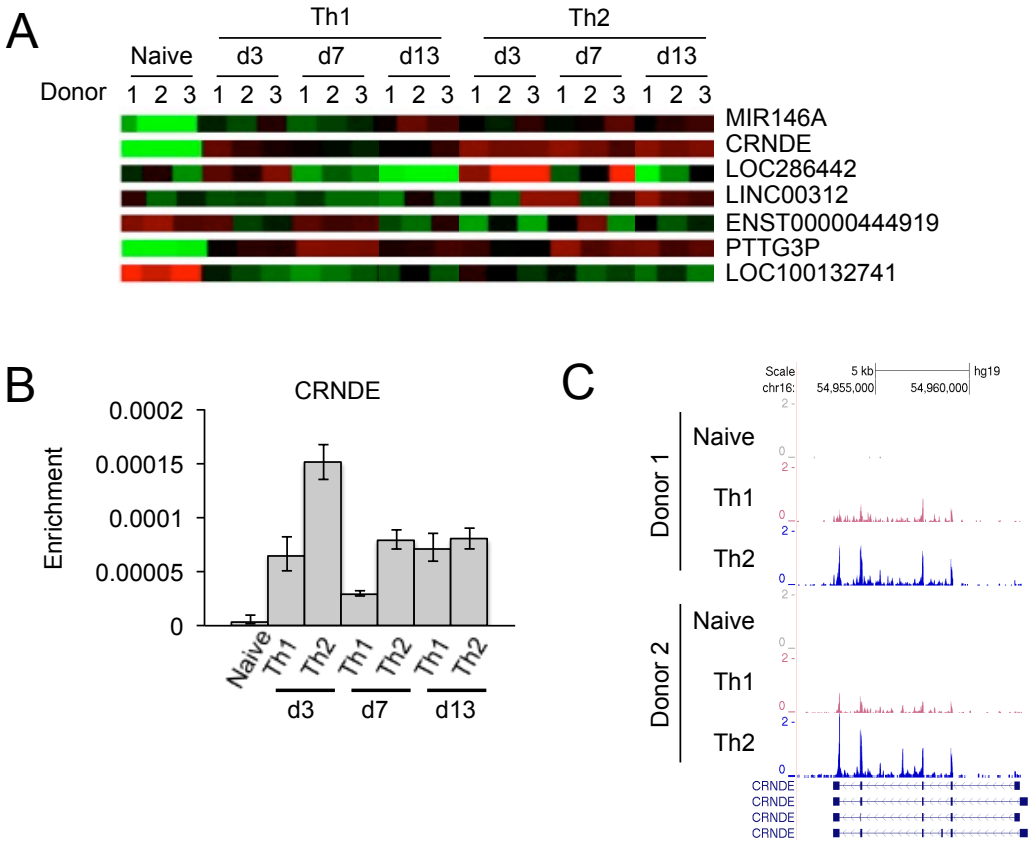


Figure 5.3: CRNDE is upregulated during T cell activation.
Expression of CRNDE was examined in unactivated naïve T cells in comparison to activated and in vitro polarised Th1 and Th2 cells. A: Microarray data comparing expression of CRNDE and other Treg ncRNAs. ENST00000415387 is not shown as it was filtered out during median centering of array data. LOC100132741 was not qPCR validated as a Treg ncRNA in Section 3.2.2 due to its high repetitive sequence but is included here as its upregulation in unactivated naïve T cells was striking. B: Quantitative PCR confirming the differential expression of CRNDE between naïve cells and *in vitro* polarization cells. C: Examination of this trend using RNA-Seq data.

5.2.3 Isoforms of CRNDE present in primary human Treg

There are 10 annotated isoforms of CRNDE (Figure 3.7) and biological associations have been made with particular isoforms. For example, gVCIn4 transcripts that contain a genomic region that is conserved among vertebrates are associated with promotion of growth and suppression of apoptosis in fibroblasts [526]. Such unspliced transcripts are downregulated by insulin signalling and appear to be antagonistic to processes downstream of insulin signalling such as glucose and lipid metabolism, while spliced transcripts are not [550]. Such unspliced transcripts may represent nascent RNA which is still tethered to DNA and therefore in close proximity to transcriptional machinery with potential influence on transcriptional regulation. As the alternative splicing of CRNDE appears to have biological consequences, and as information on the exons included would inform the design of siRNA for knockdown, I designed isoform specific primers to determine which transcript was expressed in primary human Treg. Figure 5.4 illustrates all 10 isoforms in relation to the regions detected by the microarray probe and qPCR primers and the relative enrichment of these amplicons in primary human Treg and Tresp. As this shows, the microarray probe and the original qPCR primers used for validation of microarray data (amplicon 1 [1F/R]) both detect exon 6 of CRNDE and are therefore not informative of intron inclusion. qPCR primers to potential intronic unspliced transcripts, including exon 6 to intron 6 (amplicon 2 [2F/R]) and gVCIn4 (amplicon 3 [3F/R]) show no detectable transcript by gel electrophoresis (Figure 5.4.B) or qPCR (Figure 5.4.C), while a spliced isoform including exons 2 and 6 is present (Figure 5.4.B, amplicon 4 [4F/R]) and is more abundant in Treg compared to Tresp

(Figure 5.4.C). The amplicon representing this spliced product was expected to be 349bp if it includes the longest exon 5 annotation but excludes exon 3. Gel electrophoresis confirmed an amplicon of this size in Treg (Figure 5.4.B). This supports the presence of isoforms ENST00000559598 and ENST00000502066 (Figure 5.4.A). Products at 266bp and 277bp, which reflect ENST00000501177 and ENST00000559432, respectively, are not present (Figure 5.4.A, B). Could comment on whether these results are consistent with RNA-seq data. Therefore *CRNDE* undergoes splicing in primary human Treg and spliced transcripts contribute to the differential expression seen between Treg and Tresp.

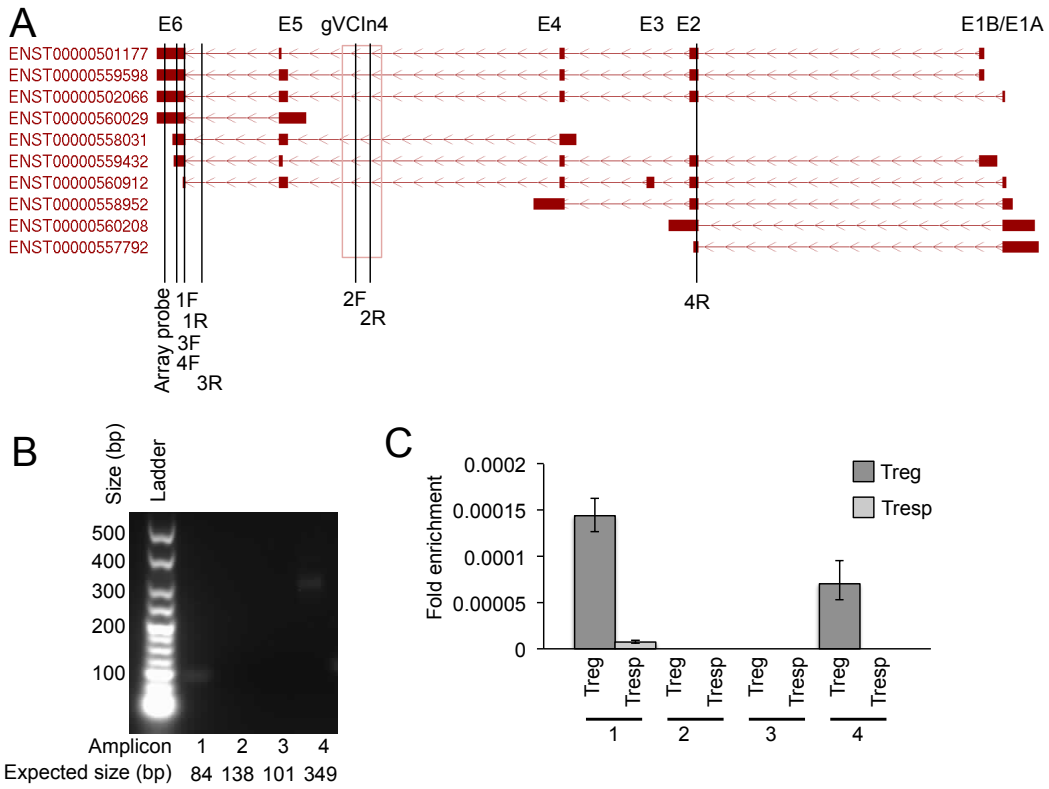


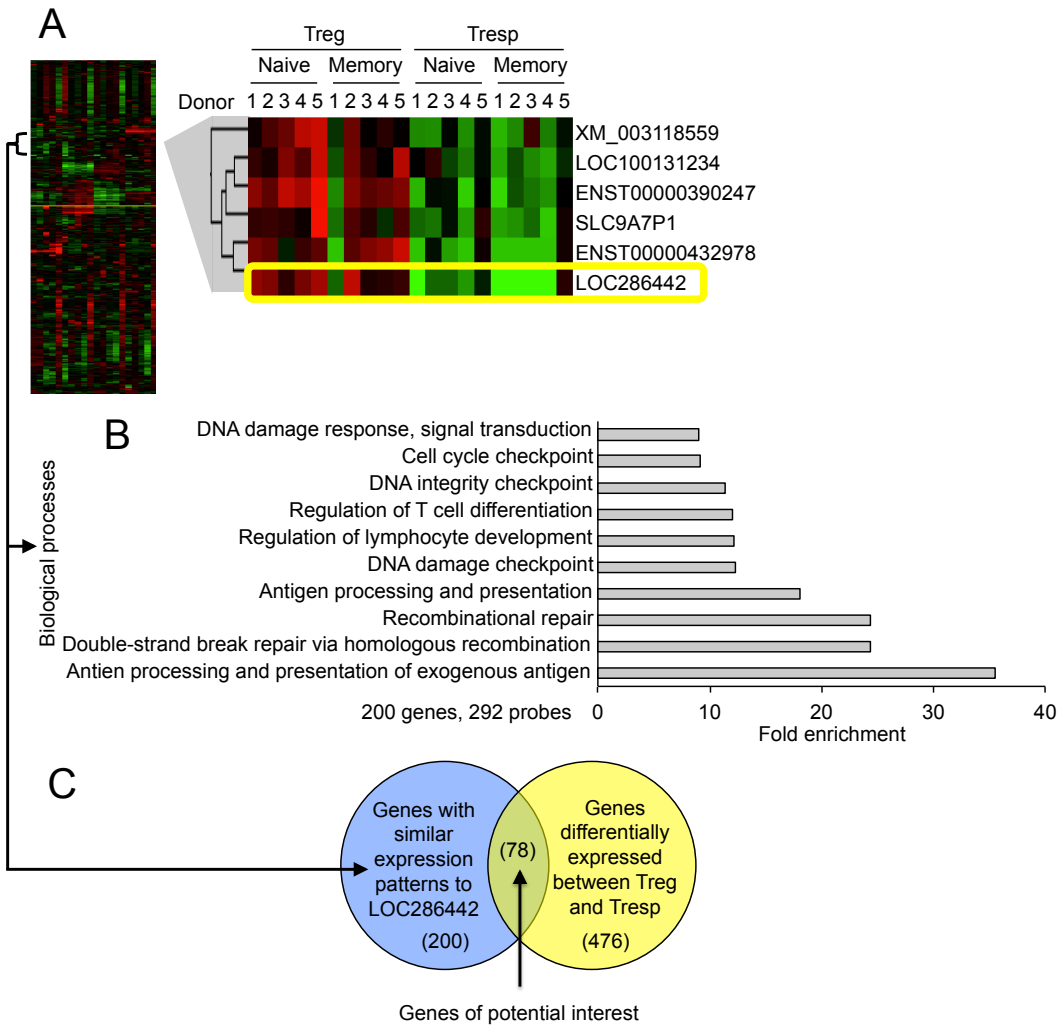
Figure 5.4: Isoforms of CRNDE expressed in human primary Treg

A: The 10 annotated isoforms of CRNDE. The approximate position of gVCIn4 described by Ellis et al. 2014 is shown by the red square. Locations of the regions detected by the microarray probe and qPCR primers are indicated. 1F and 1R correspond to the primers used for qPCR validation of CRNDE expression (Figure 3.4). B: The amplicons from the qPCR primers described in A in primary human Treg. C: The relative enrichment of each amplicon in Treg and Tresp. F: forward primer; R: reverse primer. E: exon.

5.2.4 Genes and biological processes associated with LOC286442

In order to gain insight into the role of LOC286442 in Treg, I took the same approach as described in Section 5.2.1 to identify biological processes with potential relevance to LOC286442 function. I identified genes with known functions that exhibited expression patterns similar to *LOC286442* by hierarchical clustering of microarray data (Section 2.5.2). As Figure 5.5.A illustrates, the majority of genes with the most similar expression were identifiable by accession numbers rather than HGNC gene symbols, implying their functions are not well understood. In accordance with this, literature surrounding these genes was limited and so their expression does not inform LOC286552 function. A larger cluster of genes allowed identification of statistically significantly associated biological processes through gene ontology, as shown in Figure 5.5.B (200 genes, P (adjusted) <0.05). These biological processes included T cell specific processes such as antigen processing and presentation and regulation of lymphocyte development and T cell differentiation. To select for genes with most relevance to Treg biology, I identified genes in this cluster that were also differentially expressed between Treg and Tresp (Figure 3.2.A, Table 7.1 and Table 7.2). From this list of genes (Table 7.32), several genes had notable relevance to Treg biology not least *FOXP3*, the Treg lineage-specifying transcription factor [37-39]. Also present in this list was Helios (IKZF2) a transcription factor that shows Treg-specific expression within the peripheral lymphocyte compartment. Within Treg, the expression of Helios has been proposed to be exclusive to thymically derived, 'naturally-occurring' nTreg, as opposed to peripherally induced Treg [89], however

this assertion has proved controversial [90, 91, 587] and the exact influence that Helios has over Treg function is not understood. In addition, expression of *TNFR2* (*TNFRSF1B*) was also seen in association with *LOC286442*. TNFR2 is a TNF receptor, which has been frequently associated with Treg function [217-219]. For example, in both humans and mice TNFR2 ligation leads to Treg expansion [217, 220] and expression of TNFR2 identifies Treg with increased function [220] and through a potential negative feedback loop for the control of inflammation, TNF signalling has been shown to stabilise Treg phenotype through TNFR2 [217-219]. I therefore hypothesised that LOC286442 may have a role is these processes through regulating the expression, directly or indirectly of *FOXP3*, *IKZF2* (Helios) and/or *TNFRSF1B* (TNFR2).



Gene	*Relevance*	*References*
TNFRSF1B (TNFR2)	Expression and downstream signalling is associated with enhanced Treg function	[37-39]
FOXP3	Treg lineage-specifying transcription factor	[89, 93, 217]
IKZF2 (Helios)	Transcription factor associated with nTreg compared to iTreg	[219, 220]

Figure 5.5: Gene and biological processes associated with *LOC286442* expression
Genes with similar expression patterns to *LOC286442* as identified by hierarchical clustering were used in a 'guilt-by-association' approach to identify biological processes other genes with possible relevance to LOC286442 function. A: The genes with most similar expression to *LCO286442*. B: A larger cluster of genes were used in gene ontology to identify biological processes associated with *LOC286442* expression (200 genes, P (adjusted) <0.05). C: These genes were compared with those differentially expressed between Treg and Tresp (476 genes, Figure 3.2.A, Table 7.1Table 7.2) to identify those with those most relevant to Treg biology. From this list (Table 7.32) FOXP3, IKZF2 (Helios) and TNFRSF1B (TNFR2) were recognised as particularly interesting. SLC9A7P1: solute carrier family, subfamily A member 7 pseudogene 1.

5.2.5 siRNA transfection results in prohibitively low cell viability

To assess the functional effects of CRNDE and LOC286442 ncRNAs in Treg biology, I sought to knock these transcripts down and assess the effect of their loss on phenotype and function. As described in Sections 5.2.1 and 5.2.4, CRNDE expression was associated with expression of TNFRSF8 (CD30), TNFRSF9 (4-1BB) and MKI67 (ki67) while LOC286442 expression was associated with TNFRSF1B (TNFR2), FOXP3 and IKZF2 (Helios), genes that have relevance to Treg function. Therefore, I hypothesised that CRNDE and LOC286442 may contribute to Treg function through direct or indirect influence on expression of their respective associated genes. In order to test this, I designed siRNAs against CRNDE and LOC286442 transcripts to assess the effect of loss of these ncRNAs on expression of these genes.

As described in Section 5.1.1 and Figure 5.1, several strategies exist for knockdown of transcripts. As ncRNAs may function within the nucleus to influence gene expression through regulation of transcription, and as shRNA mediated gene expression requires cellular components restricted to the cytoplasm, I attempted knockdown using siRNA. In support of this approach, siRNA is commonly used for knockdown of non-coding transcripts [449, 456, 588], while publications involving shRNA for this purpose are limited. I therefore used a nucleofection procedure recommended for unstimulated human T cells (Section 2.2.10.1 [469]). Assuming transfection would not be efficient for 100% of cells, a fluorescent indicator (siGLO) was co-transfected with target siRNA (in a

1:1 ratio) to allow selection for transfection competent cells by FACS. In addition, a greater number of cells per transfection condition, with a minimum of $7x10^5$ but an ideal range of $5\text{-}10x10^6$, were recommended for increased viability. National Blood Service buffy coats were used to obtain these numbers of Tregs, which contribute 1% of PBMCs, and yielded around $10x10^6$ Treg per donor allowing for 3 conditions (knockdown of FOXP3 and a Treg ncRNA alongside a non-targeting control) with $2x10^6$s per condition. siRNA to Treg ncRNAs were designed using an online algorithm recommended by the manufacturer (Section 2.2.10.1, [467]) and this was compared with an off-the-shelf siRNA to FOXP3. Knockdown was assessed in comparison to a non-targeting control. As Figure 5.6.A shows, siRNA transfection resulted in low cell viability compared to an untransfected control, with 13.7% live cells after transfection, as indicated by lack of DAPI staining. This is in comparison with the manufacturer's suggestion of 80% viability [469]. The transfection efficiency of DAPI-negative cells was 79.7% (Figure 5.6.B), which is much higher than the manufacturer's suggestion of 40% [469], possibly indicating that the use of the siGLO transfection indicator does not accurately reflect transfection competency. I sorted the live population into siGLO-positive and negative fractions and, as the yield of cells was too low to perform functional assays (Figure 5.6.C), assessed the knockdown efficiency of the commercially available FOXP3 siRNA compared to the siRNA I had designed to Treg ncRNA LOC286442. As Figure 5.6.D shows, the siRNA against FOXP3 resulted in a 31% decrease in FOXP3 RNA compared to the non-targeting control. However, the high standard error, comparative enrichment of FOXP3 RNA in siGLO- controls and low cell viability limited the reliability of

this observation. No knockdown was seen using the siRNA targeted at LOC286442.

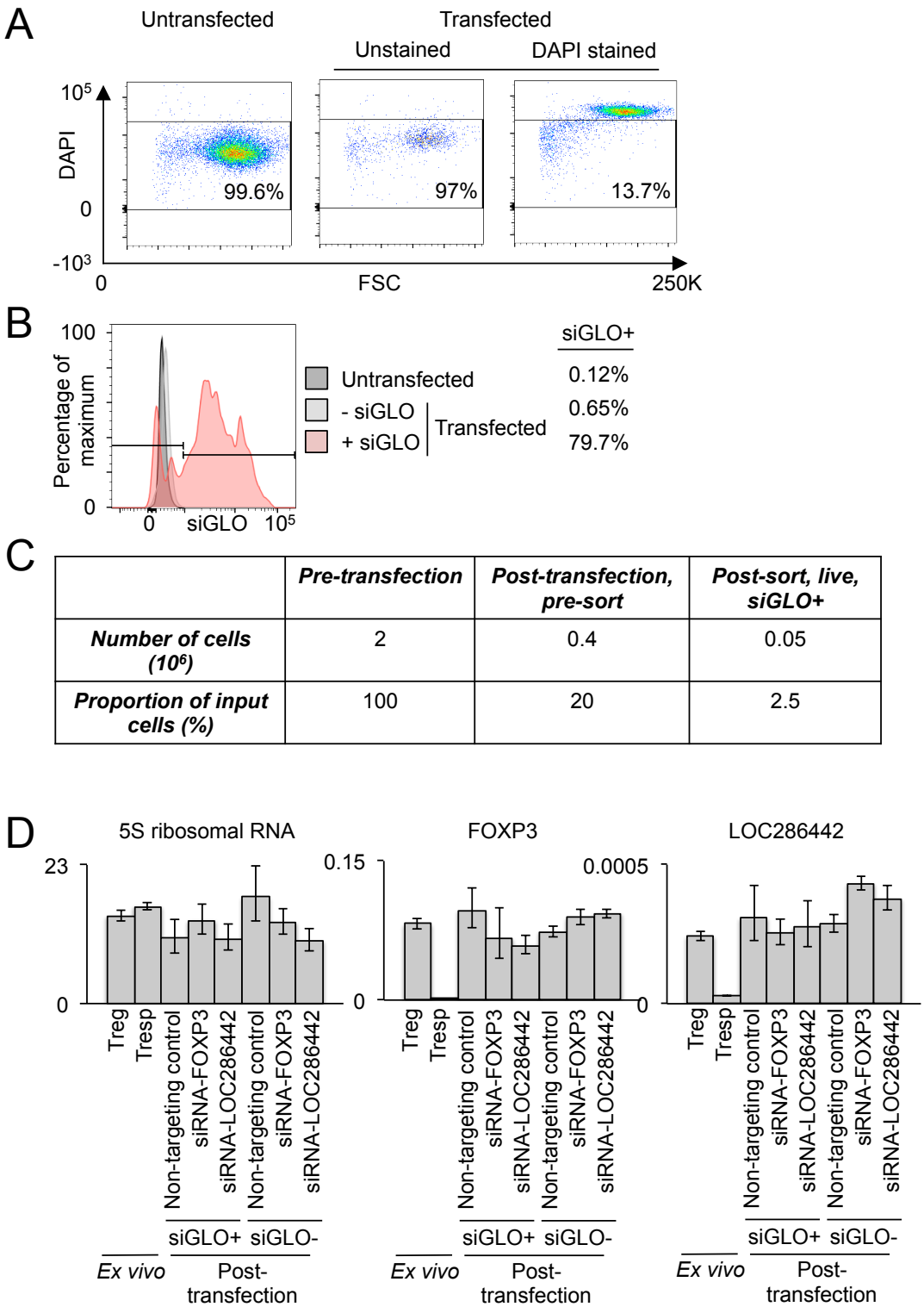


Figure 5.6: siRNA transfection results in prohibitively low cell viability
In order to investigate the functional role of Treg ncRNAs, knockdown was attempted through transfection of siRNA A: Cell viability measured by DAPI staining. B: Transfection efficiency as measured with the siGLO transfection indication. C: Comparison of numbers of cells before and after transfection and sorting for the live, transfection competent population. D: Knockdown efficiency of siRNA targeted against FOXP3 and the Treg ncRNA LOC286442. 5S ribosomal RNA is shown as a housekeeping control. siGLO- cells were also collected as a negative control..

5.2.6 Increased cell number through Treg expansion does not aid knockdown

The low cell viability post-transfection as shown in Section 5.2.5 (Figure 5.6) may have related to the low number of cells at the point of transfection: Although $2x10^6$ cells were used, which is above the minimum recommendation of $7x10^5$, the ideal range is $5\text{-}10x10^6$ cells [469]. The use of primary human Treg, which at 1% of PBMCs are infrequent in peripheral blood, limited the number of cells per condition and so I attempted to increase cell number through *in vitro* expansion of Treg before performing transfections, as described in Section 2.2.8 For the use of stimulated cells $1\text{-}5x10^6$ cells per transfection was recommended alongside the use of alternative electroporation programs (Section 2.2.10.1, [469]). I therefore compared these recommended Amaxa programs and using various cell numbers. As Table 5.1 demonstrates, increasing the number of cells and altering the transfection program used did not improve cell viability.

Table 5.1: Changes in Treg number and transfection programs do not enhance the efficiency of siRNA transfection

Transfection program	*Number of cells (10^6)*	*Proportion of live cells (% of total)*	*Proportion of siGLO+ cells (% of live)*
U-014	2.5	7.3	92.1
	5.0	8.0	82.6
T-023	2.5	11.6	25.3
	5.0	13.9	31.9
T-020	2.5	9.2	28.9
	5.0	9.8	22.7

5.2.7 Electroporation with Agile Pulse shows improved cell viability but siRNA promotes death in T cells

A novel electroporation technology recently became available and has been successfully applied to the transfection of primary human T cells (personal correspondence with Laurie Menger, [589]). The Agile Pulse (BTX® Harvard Apparatus) differs from conventional electroporation and Amaxa nucleofection through variation of the combination of pulse voltage, pulse duration and interval duration during the sequences of pulses applied. This allows, for example, a combination of initial short but high voltage pulses followed by lower voltage pulses of a longer duration. This has been shown to increase transfection efficiency [589] and combinations of pulses have been optimised for successful transfection of T cells with mRNA for TALEN-mediated genome editing (Laurie Menger, Karl Peggs, personal communication). Using a GFP mRNA as a comparator with known capability to transfect T cells, I examined the use of the Agile Pulse for transfection of siRNA into T cells. As Figure 5.7 shows, 87.3% of cells were viable 24hr post-transfection with GFP mRNA (x M). However, an equimolar concentration of siRNA resulted in only 13% live cells, a 6.7 fold decrease in viability compared to GFP mRNA. Titrating the concentration of siRNA showed that cell viability was proportional to siRNA concentration, demonstrating that siRNA transfection is toxic to T cells. In addition, siRNA concentration was proportional to transfection efficiency, with the lowest molarity resulting in 23% of transfected cells, compared to 88% for the highest molarity. Therefore, the use of siRNA requires a compromise between cell viability and transfection efficiency. The toxic effect of siRNA compared to mRNA may relate to the similarity of double

stranded siRNA to viral epitopes, which may be recognised by TLR3 on T cells leading to apoptosis [590].

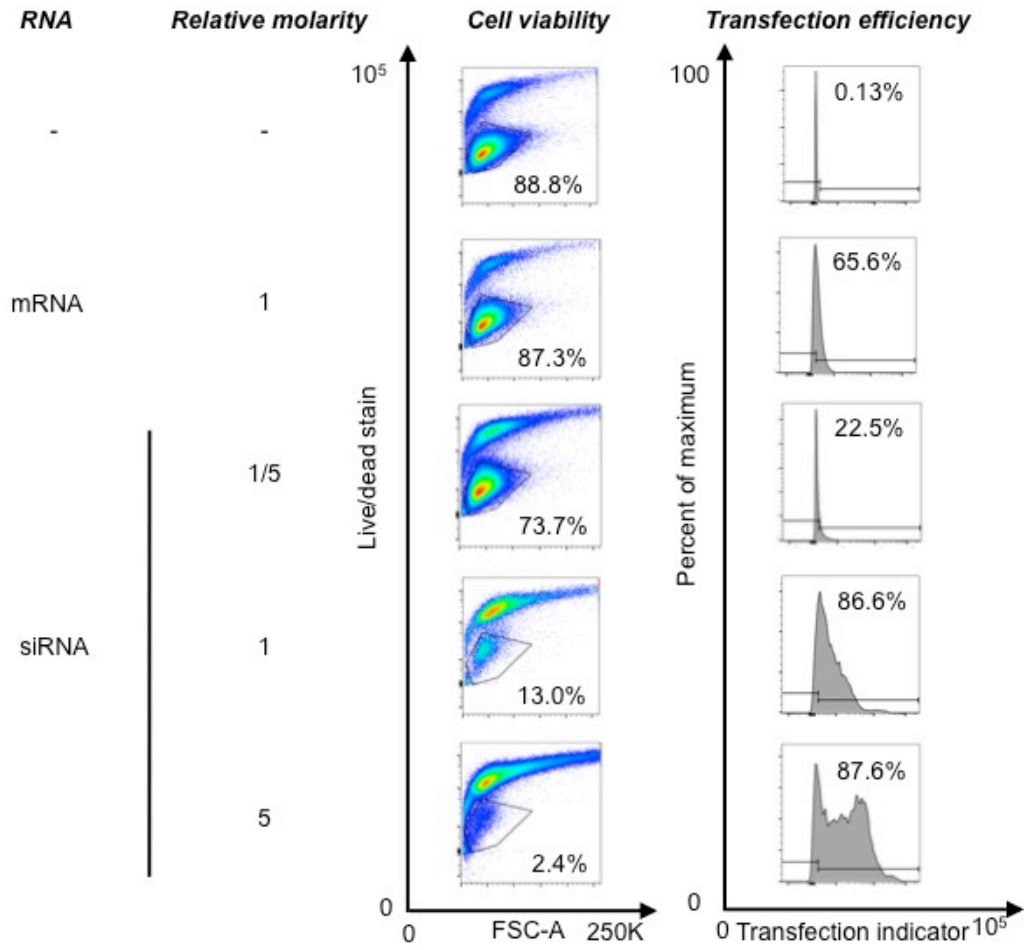


Figure 5.7: Transfection of siRNA leads to decrease in cell viability

The viability and transfection efficiency of transfection using Agile Pulse electroporation was compared for transfection of mRNA (for GFP) and an equimolar concentration of siRNA. The concentration of siRNA was titrated to examine the direct effect of siRNA on viability. .

5.2.8 Locked nucleic acid treatment achieves effective knockdown in T cells

Due to the lack of success using siRNA I attempted instead to knockdown ncRNAs using LNAs (Section 5.1.1) LNAs against CRNDE and GAPDH were designed in collaboration with RaNA Therapeutics and LNAs against LOC286442 were designed by Exiqon. I treated T cells with a fluorescently tagged non-targeting control LNA as described in Section 2.2.11 and assessed cell viability, LNA uptake and knockdown efficiency. As Figure 5.8.A shows, 90.7% of cells were viable after LNA treatment, which was comparable with cells not treated with LNA, revealing no obvious toxic effects of LNA unlike that seen with siRNA (Figure 5.7). Titration of LNA targeted towards GAPDH (Figure 5.8.C), demonstrated a dosage effect with lower transcript level detected with increasing concentration of LNA. This effect was specific to GAPDH; although transcript levels were variable with other conditions, the same dose response was not seen for the non-targeting control not for 5S ribosomal RNA.

Surprisingly, 99.81% of cells exhibited fluorescence after incubation with non-targeting control LNA, suggesting that nearly all cells had taken it up (Figure 5.8.B). To confirm such surprisingly highly efficient uptake, I hypothesised that this high proportion of LNA-positive cells could potentially be explained by adherence of LNA to the cell surface rather than actual entry into the cell. In this circumstance, any decrease in target level could be due to changes in gene expression resulting from signalling downstream of cell surface:LNA interaction rather than an LNA-RNA interaction leading to RNA degradation. For this reason, internalisation

analysis was performed through integration of flow cytometry and microscopy (Image Stream Technology, Section 2.2.4). Using markers for the cell surface (CD4) and nucleus (Hoerchst), I compared the similarity of fluorescent LNA to these locations through measurements of internalisation (comparison of LNA with CD4), nuclear localisation (comparison of LNA with Hoerchst) and co-localisation. As Figure 5.9.A shows, the median score for LNA internalisation was 1.821, which was not dissimilar to internalisation of the nucleus, a positive control, which scored 1.969. The nuclear localisation score for CD4 was 1.641 while nuclear localisation of LNA was greater with a median score of 2.645. As a negative control for co-localisation, CD4 (cell surface) and Hoerchst (nucleus) were compared, giving a similarity score of 0.7874. The co-localisation of LNA with both the cell surface and the nucleus gave greater similarity scores of 1.166 and 1.268, respectively. These measurements are therefore consistent with cellular uptake of LNA.

It was possible that LNA uptake was heterogeneous in the cell population with some cells demonstrating increased uptake compared to others. Such a bimodal population distribution could allow analysis of phenotypic effects to be restricted to cells with optimum uptake. As Figure 5.9.B shows, I examined the co-localisation of LNA with CD4 against the co-localisation of LNA with the nucleus. The cell population appeared unimodal, indicating that similar uptake of LNA occurs in all cells. Figure 5.9.C illustrates the co-localisation of fluorescence for a representative cell.

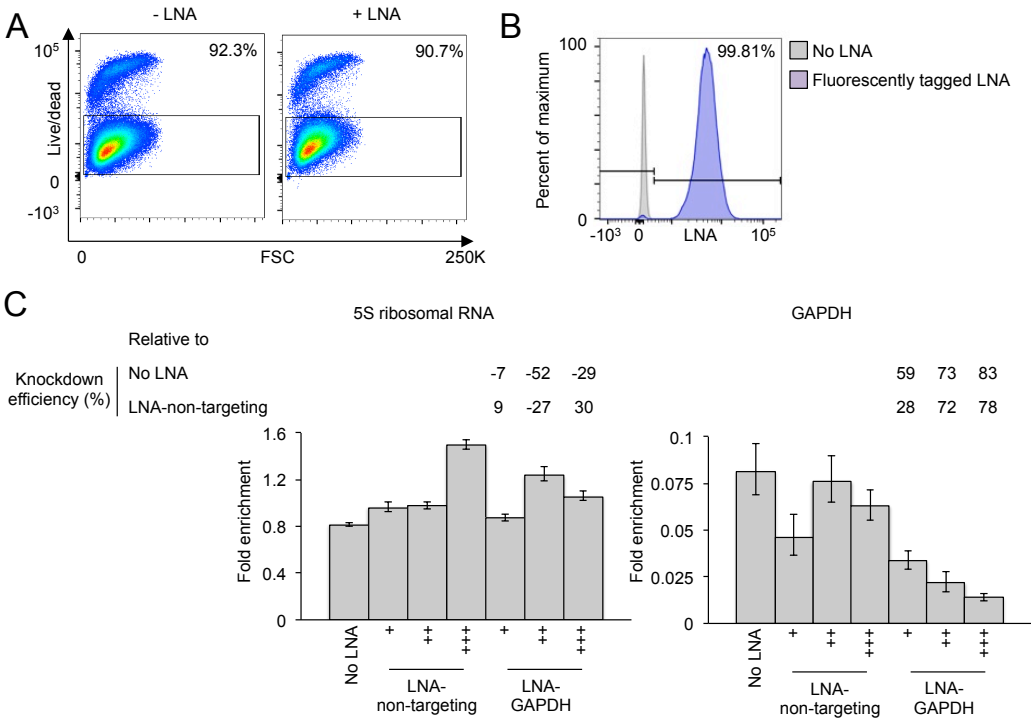


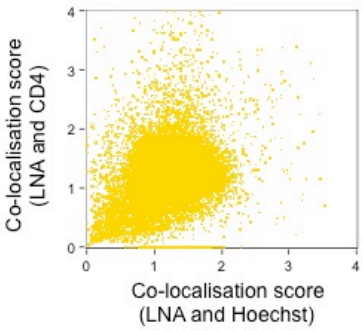
Figure 5.8: Knockdown using LNAs shows unexpected efficiency of LNA uptake by CD4 T cells.

A: Cell viability after culture with and without LNA treatment. B: Assessment of LNA uptake by CD4 T cells through gymnosis. C: Targeting of LNAs towards the GAPDH transcript. 5S ribosomal RNA is given as a control for non-specific effects of knockdown. Cultures treated with a non-targeting LNA and no LNA at all were also assessed.

A

Parameter	*Similarity score (median for population)*
Internalisation	
of nucleus (positive control)	1.969
of LNA	1.821
Nuclear localisation	
of CD4 (negative control)	1.641
of LNA	2.645
Co-localisation	
of surface with nucleus (negative control)	0.7874
of LNA with surface	1.166
of LNA with nucleus	1.268

B



C

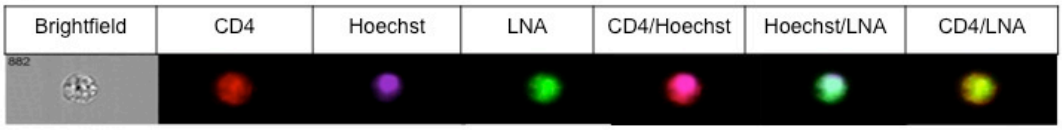


Figure 5.9: Internalisation of LNA by CD4 T cells

Internalisation of LNA by CD4 T cells was assessed by Image Stream Technology (Amnis) A: Median scores for fluorescence similarity comparing cell surface (CD4), nucleus (Hoechst) and LNA gated on focussed, single cells. Internalisation is the fluorescence similarity of a parameter with the cell surface (CD4). Nuclear localisation is the fluorescence similarity of a parameter with the nucleus (Hoechst) Co-localistion is the fluorescence similarity of two chosen parameters. Fluorescence similarity is measured as bright detail similarity and reflects the correlation coefficient between the pixel values of two image pairs. A lower score indicates less similarity in fluorescence and therefore a lower physical association. B: A unimodal population distribution of LNA localisation is seen. Co-localisation score is the ratio of bright detail similarity between two parameters. C: An example representative of the focussed, single cell population illustrating distribution of LNA through the cell and physical overlap of LNA with both the nucleus (Hoechst) and cell surface (CD4). Brightfield demonstrates cellular morphology with white light.

5.2.9 Effect of CRNDE knockdown on Treg phenotype

In order to interrogate the role of CRNDE upregulation in Treg versus Tresp, I knocked CRNDE down using LNAs, as optimised in Section 5.2.8. Figure 5.10 demonstrates that knockdown of CRNDE using LNAs was effective. Compared to no LNA-control, CRNDE LNA resulted in a 67% reduction in CRNDE transcript level and compared to non-targeting control LNA, CRNDE LNA resulted in a 71% reduction. I hypothesised that knockdown of CRNDE may affect expression of the co-regulated genes I had identified by cluster analysis (Section 5.2.1), and therefore assessed changes in their expression that correlated with CRNDE knockdown. These genes of interest included those encoding the TNF receptor superfamily members CD30 (TNFRSF8) and 4-1BB (TNFRSF9), signalling through which has been related to Treg function [217-219, 221, 583-585]. The gene for ki67 (MKI67), which is upregulated upon cell cycling, was also identified as being co-expressed with CRNDE. This association may suggest role for CRNDE in general T cell proliferation rather than a Treg specific role and with this in mind I assessed changes in ki67 in total T cells as well as Treg. In addition I assessed the impact of CRNDE knockdown on expression of the Treg lineage-specifying transcription factor FOXP3.

The phenotypic correlations of Treg and total T cells in respect to these genes of interest are presented in Figure 5.11. As shown, upon knockdown of CRNDE, changes for most proteins of interest were not significant when compared to the non-targeting control and cells cultured without LNA treatment. Expression was also similar in cells treated with an LNA

targeting GAPDH. CRNDE knockdown did however show a significant difference in the proportion of CD30 positive cells when compared to the non-targeting control (Figure 5.11.A middle, p=0.04). However, there was no difference to non-treated cells and so this could instead represent an increase in CD30 expression in the control LNA-treated population.

While performing these experiments I confirmed that donors showed the expected upregulation of CRNDE in Treg compared to Tresp. In doing this I noticed that expression of CRNDE was reduced between *ex vivo* Treg and those activated *in vitro* during LNA knockdown experiments (Figure 5.12). This was unexpected given the association of CRNDE with activation (Section 5.2.2). Indeed, consistent with the array data, expression in Tresp increased with *in vitro* activation (Figure 2.12). This suggested that CRNDE expression is differentially regulated in Treg compared to Tresp.

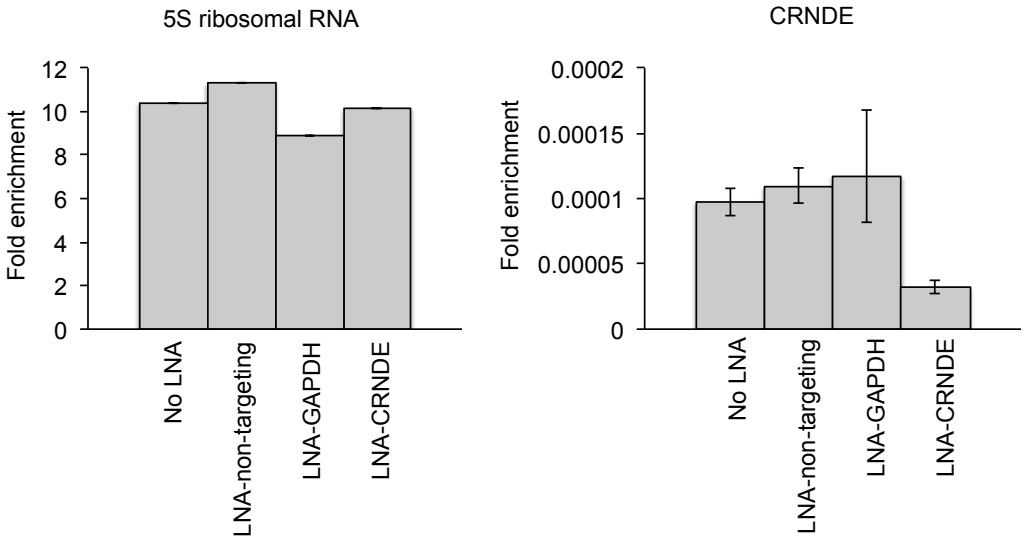


Figure 5.10: Knockdown of CRNDE

Knockdown of CRNDE using LNAs targeted against the transcript was assessed by qPCR in comparison to treatment without LNAs and treatment with a non-targeting control. Non-specific decrease in transcript level was assessed by comparison with an LNA targeted at the GAPDH transcript. Fold enrichment is relative to HPRT1.

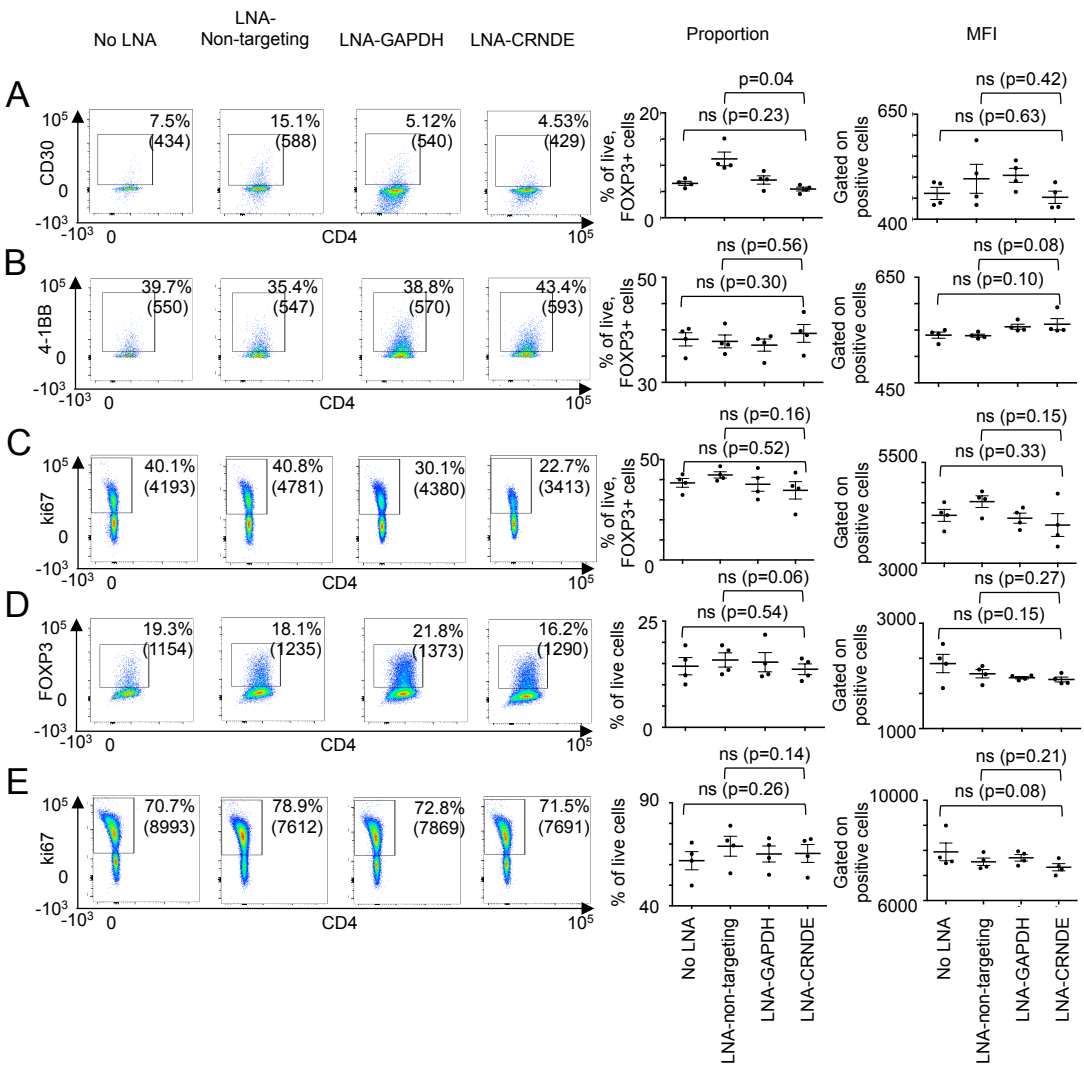


Figure 5.11: Effect of CRNDE knockdown on expression of CD30, 4-1BB, ki67 and FOXP3.

Genes correlating with CRNDE expression were assessed by flow cytometry for changes that correlated with CRNDE knockdown. A: Changes in CD30 expression upon treatment with LNAs targeted against CRNDE and GAPDH as well as a non-targeting LNA and no LNA at all. Representative FACS plots gated on FOXP3+ cells are shown along with percentage of CD30-positive cells and MFI of these cells indicated in brackets (left). Data from four biological replicates for the proportion of CD30-positive cells (middle) and MFI (right). B: As for A, except for 4-1BB. C: as for A except ki67. D: As for A except for FOXP3 and for the whole population. E: As for D except for ki67. Plots show mean and SEM. Significance was estimated using two-tailed, paired t test.

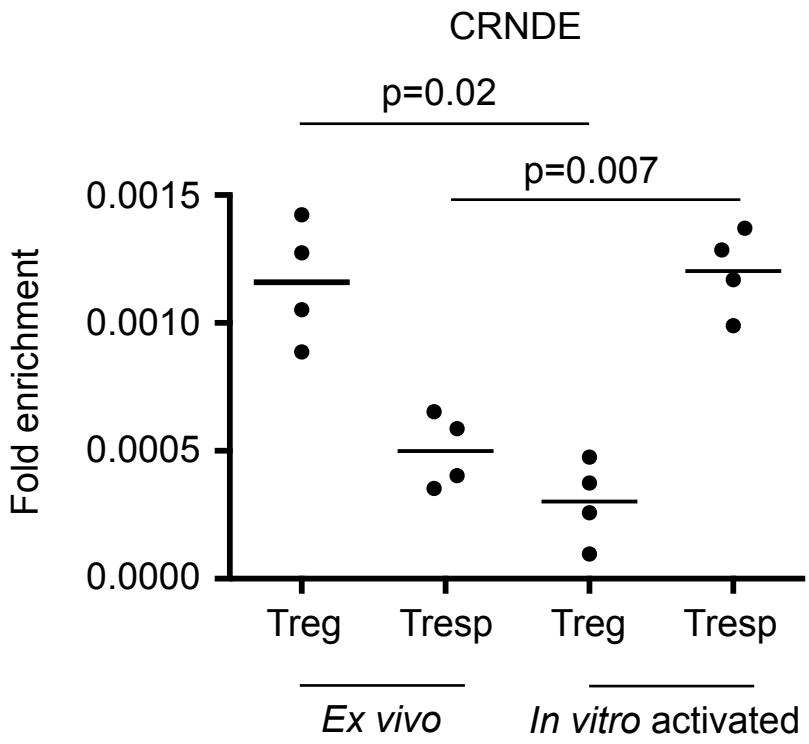


Figure 5.12: Differential regulation of CRNDE in Treg and Tresp
Comparison of the level of CRNDE expression in Treg and Tresp *ex vivo* and upon *in vitro* activation. Fold enrichment is relative to HPRT1. Plots show mean and SEM. Significance was estimated using two-tailed, paired t test.

5.2.10 Effect of LOC286442 knockdown on Treg phenotype

I also performed knockdown experiments for the Treg ncRNA LOC286442 in order to gain insight into its functional role in Treg biology. Again, knockdown was attempted using LNAs and Figure 5.13 demonstrates the efficiency of this approach for *LOC286442*. Compared to the non-targeting control LNA, there was a 40% reduction in transcript level in cells treated with an LNA to LOC286442 and compared to culture without LNA, there was a 34% reduction. As for CRNDE (Section 5.2.9), I assessed the effect of this knockdown on the expression of Treg genes with similar expression patterns to *LOC286442*; FOXP3, Helios and TNFR2. Figure 5.14 presents the effect of LNA treatment on the expression of these proteins demonstrating that no significant difference could be seen.

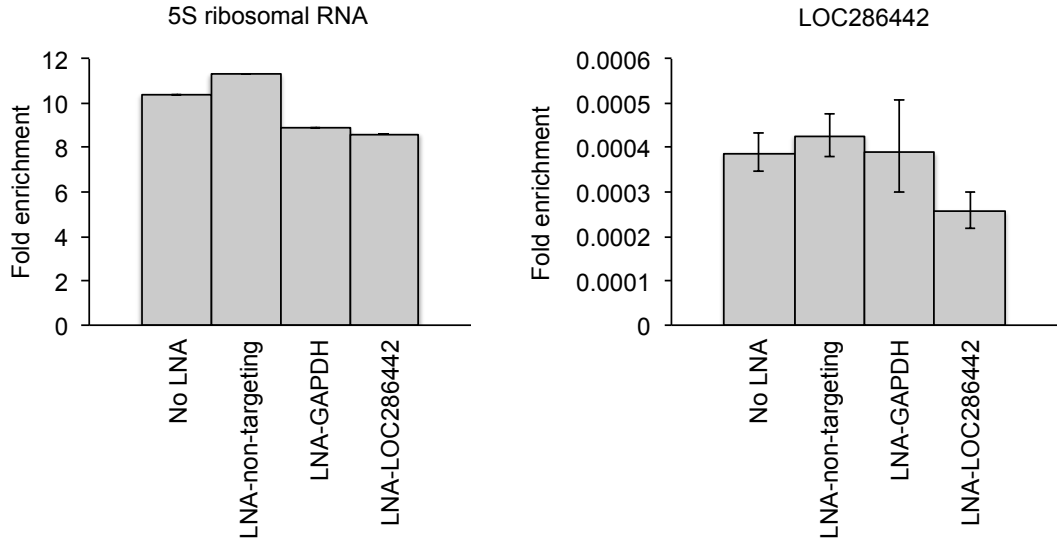


Figure 5.13: Knockdown of LOC286442

Knockdown of LOC286442 using LNAs targeted against the transcript was assessed by qPCR in comparison to treatment without LNAs and treatment with a non-targeting control. Non-specific decrease in transcript level was assessed by comparison with an LNA targeted at the GAPDH transcript. Fold enrichment is relative to HPRT1

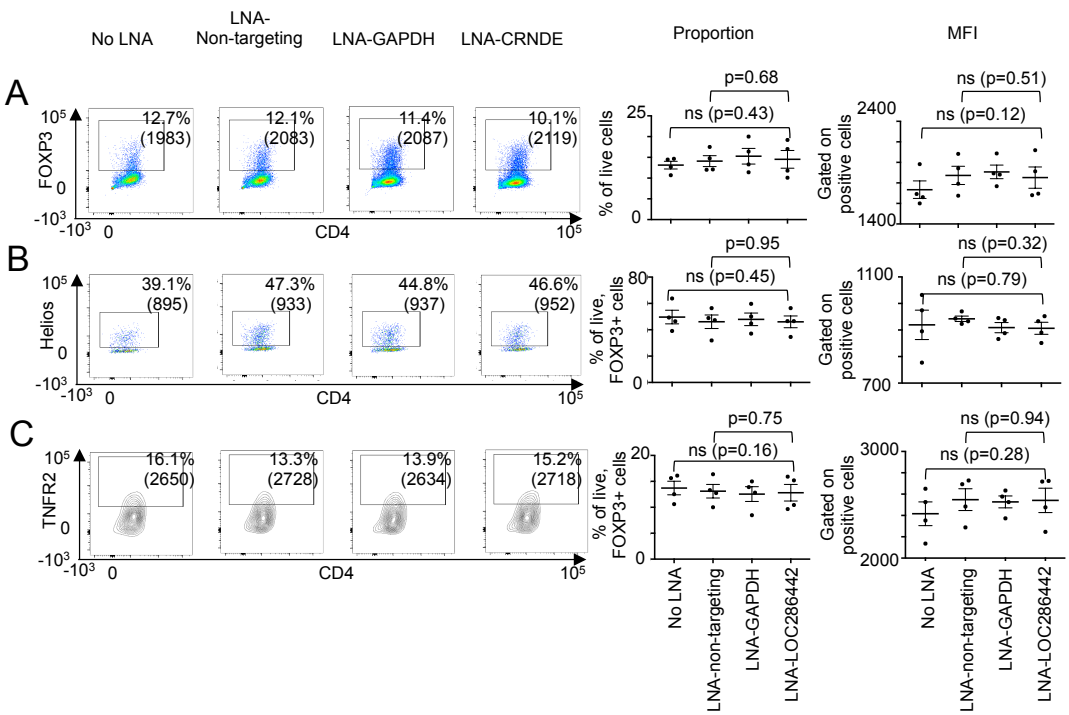


Figure 5.14: Effect of LOC286442 knockdown on expression of FOXP3, Helios and TNFR2.

Genes correlating with LOC296442 expression were assessed by flow cytometry for changes that correlated with LOC286442 knockdown. A: Changes in FOXP3 expression upon treatment with LNAs targeted against LOC286442 and GAPDH as well as a non-targeting LNA and no LNA at all. Representative FACS plots gated on total live cells are shown (left) along with percentage of FOXP3-positive cells and MFI of these cells indicated in brackets. Data from four biological replicates is given for the proportion of FOXP3-positive cells (middle) and MFI (right). B: As for A, except for Helios and gated on FOXP3+ cells. C: as for B except TNFR2. Plots show mean and SEM. Significance was estimated using two-tailed, paired t test

5.2.11 Effect of IFNG eRNA knockdown on IFNG expression

As described in Chapter 4, I have identified eRNAs from T-bet super-enhancers upstream of Th1 specific genes. To determine if these eRNAs could play a role in *IFNG* activation, I sought to knockdown eRNAs from the *IFNG* super-enhancer using LNAs. I therefore polarised Th1 cells *in vitro* (Section 2.2.9) in the presence of LNA to IFNG eRNAs, GAPDH or a fluorescent non-targeting control (Section 2.2.11). As Figure 5.15 shows, cells were successfully polarised, as represented by IFNγ expression (Figure 5.15.A) and LNA uptake-efficiency was similar to that seen in the optimisation experiments described in Section 5.2.8. Th1 polarised cells were treated with 8 different LNAs separately and also pooled together. These LNAs were targeted at 4 different RNA-Seq peaks at the *IFNG* super-enhancer, with each peak being targeted by 2 different LNAs. Figure 5.15.C describes the proportion of IFNγ positive cells for each LNA treatment and the median fluorescence intensity (MFI) of the IFNγ positive population. The fold difference for both the proportion and MFI are given relative to the non-targeting control. For donor 1, both LNAs targeting eRNA4 resulted in a change in IFNγ expression compared to the non-targeting control with a modest increase in the proportion of IFNγ positive cells (1.2 fold for both LNA-7 and LNA-8) but around half the MFI (as shown by a fold difference of 0.5 and 0.6 for LNA-7 and LNA-8, respectively). The same trend was also seen for the pool of LNAs, with an increase in the proportion of IFNγ cells by 1.3 fold and a 0.5 fold difference in MFI). For donor 2 this trend was not observed with treatment of LNAs targeted at eRNA4, but a similar observation was noted for both LNAs targeted at eRNA3. In this circumstance the proportion of IFNγ positive

cells was similar to the non-targeting control (as indicated by a fold difference of 1.0 and 1.1 for LNA-5 an LNA-6, respectively) but the MFI was again reduced by around a half (as shown by a fold difference of 0.4 and 0.5 for LNA-5 and LNA-6, respectively). For both donors, LNA-2, which was targeted at eRNA1 showed a 1.2-fold increase in the proportion of IFNγ positive cells. In addition, LNA targeted to GAPDH, which does not have a direct effect on *IFNG* transcription, showed at similar increase in the proportion of IFNγ positive cells as LNAs for eRNA3 in donor 1 (1.2 fold), along with a decrease in MFI (0.7 fold). In donor 2, an increase in MFI was seen under this condition.

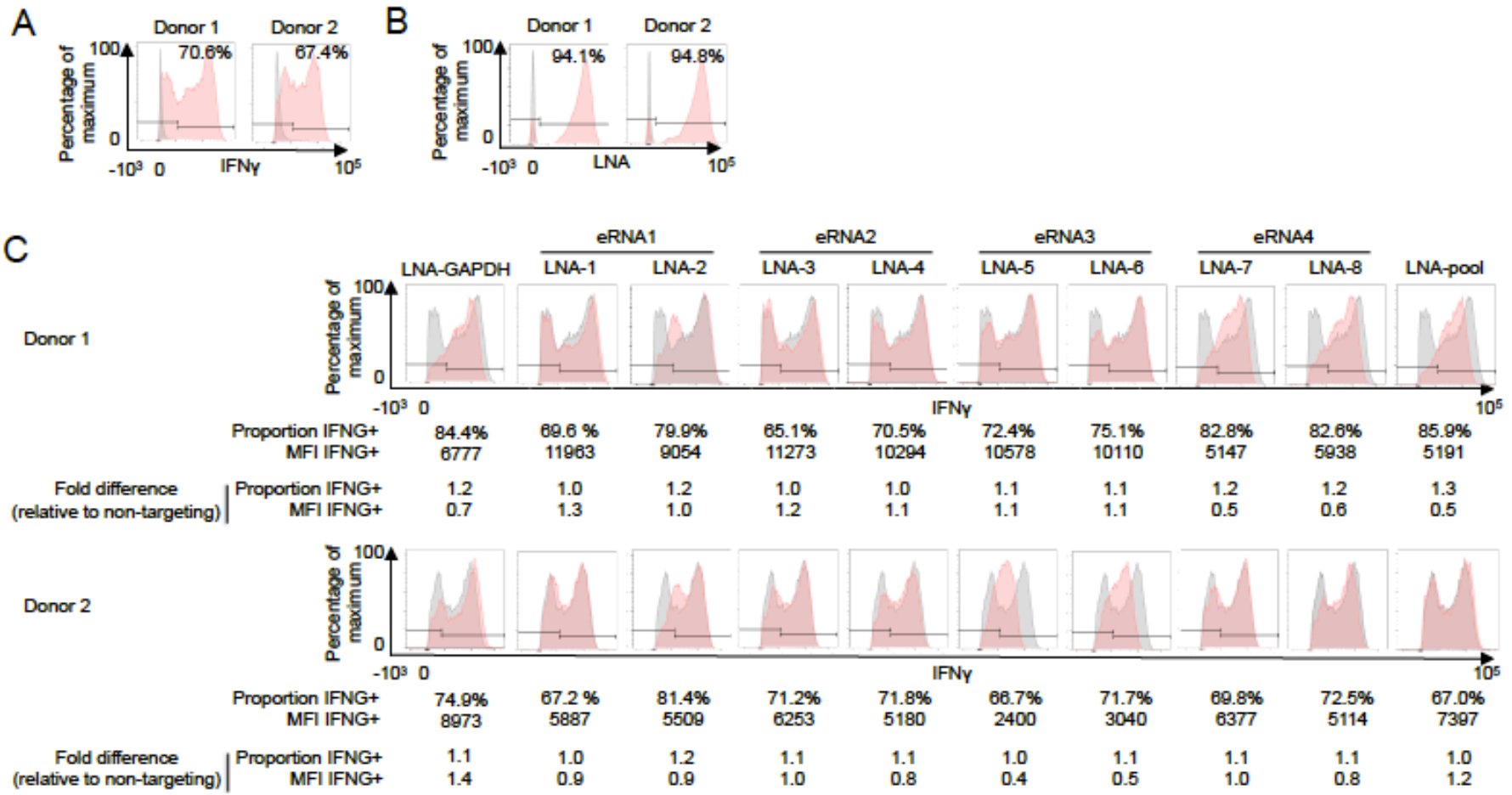

A
Donor 1
Donor 2
70.6%
67.4%
Percentage of maximum
100
0
-10³ 0
IFNγ
10⁵
B
Donor 1
Donor 2
94.1%
94.8%
Percentage of maximum
100
0
-10³ 0
LNA
10⁵
C
eRNA1
eRNA2
eRNA3
eRNA4
LNA-GAPDH
LNA-1
LNA-2
LNA-3
LNA-4
LNA-5
LNA-6
LNA-7
LNA-8
LNA-pool
Donor 1
Percentage of maximum
100
0
-10³ 0
IFNγ
10⁵
Proportion IFNG+ 84.4% 69.6 % 79.9% 65.1% 70.5% 72.4% 75.1% 82.8% 82.6% 85.9%
MFI IFNG+ 6777 11963 9054 11273 10294 10578 10110 5147 5938 5191
Fold difference (relative to non-targeting)
Proportion IFNG+ 1.2 1.0 1.2 1.0 1.0 1.1 1.1 1.2 1.2 1.3
MFI IFNG+ 0.7 1.3 1.0 1.2 1.1 1.1 1.1 0.5 0.6 0.5
Donor 2
Percentage of maximum
100
0
-10³ 0
IFNγ
10⁵
Proportion IFNG+ 74.9% 67.2 % 81.4% 71.2% 71.8% 66.7% 71.7% 69.8% 72.5% 67.0%
MFI IFNG+ 8973 5887 5509 6253 5180 2400 3040 6377 5114 7397
Fold difference (relative to non-targeting)
Proportion IFNG+ 1.1 1.0 1.2 1.1 1.1 1.0 1.1 1.1 1.1 1.0
MFI IFNG+ 1.4 0.9 0.9 1.0 0.8 0.4 0.5 1.0 0.8 1.2

Figure 5.15:Effect of *IFNG* eRNA knockdown on *IFNG* expression

(From previous page) Th1 polarised cells from two donors were treated with LNAs targeted to eRNAs transcribed from the IFNG super-enhancer and the effect on IFNγ level assessed by flow cytometry. A: Polarisation of Th1 cells as shown by expression of IFNγ in stimulated cells not treated with LNA relative (red) to an unstimulated control (grey). B: LNA efficiency demonstrated with a fluorescently tagged LNA (red) relative to an untagged LNA (grey). C: Th1 polarised cells were treated with 8 different LNAs separately and pooled together. These LNAs were targeted at 4 different RNA-Seq peaks at the IFNG super-enhancer with each peak being targeted by 2 different LNAs. The proportion of IFNγ positive cells in LNA-eRNA treated cells are given, as well as the median fluorescence intensity of the IFNγ positive population. The fold difference relative to a non-targeting control is also provided. MFI: median fluorescent intensity.

To assess if these differences in IFNγ expression by flow cytometry correlated with a specific decrease in the level of eRNA transcript targeted by each LNA, I performed qPCR (Figure 5.16.A). As eRNAs are expected to influence gene expression at the level of transcription I also examined the effect of eRNA knockdown on the level of IFNG mRNA by qPCR (Figure 5.16.B). As Figure 5.16 shows *IFNG* RNA increased in abundance upon stimulation in both donors and, in donor 1, eRNA3 and eRNA4 were both increased upon stimulation in parallel with this. Although both eRNAs were detectable in donor 2, the increase in abundance upon stimulation was not observed. For both donors, there was no appreciable difference in expression of *IFNG* or either eRNA when a non-targeting control or an LNA targeting GAPDH was introduced and so there is no evidence of off-target or non-specific effects through these controls. However, there was also no obvious difference in IFNG, eRNA3 or eRNA4 RNA abundance in cells treated with LNAs targeting these eRNAs versus the non-targeting control. Therefore, transcript level does not reflect the trend seen at the protein level by flow cytometry (Figure 5.15).

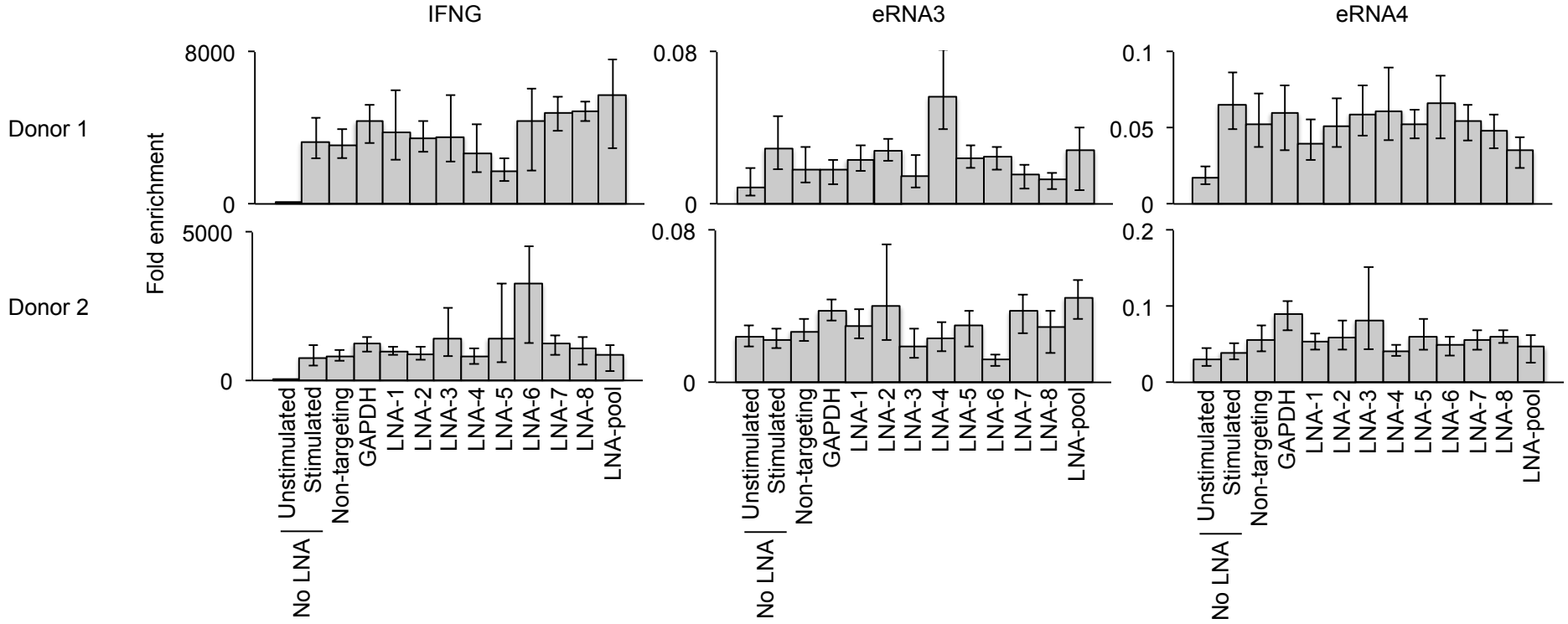


Figure 5.16: Effect of LNA of the level of targeted eRNAs from the IFNG super-enhancer and the corresponding level of IFNG mRNA transcript.

Knockdown of targeted eRNAs and IFNG transcript level was assessed by qPCR.

5.3 Discussion

In the experiments described in this chapter I have explored the functional relevance of ncRNAs identified in Chapters 0 and 4 for T cell lineage differentiation. In doing this, I have focused on the Treg lincRNAs CRNDE and LOC286442 as well as eRNAs from T-bet super-enhancers using the IFNG super-enhancer as a model.

Examining genes with associated expression patterns to the Treg lincRNAs, I observed that CRNDE was associated with biological processes such as cell cycling and proliferation (Figure 5.2). *In vitro* Th1 and Th2 polarisation data supported a role for CRNDE in these processes in T cells in general, with upregulation being seen in *in vitro* activated cells compared to *ex vivo* unactivated naïve T cells (Figure 5.3). This is also consistent with published work that has demonstrated an association between CRNDE expression and several cancers [524, 548]. Surprisingly, I observed that Treg and Tresp differed in this respect, with CRNDE level decreasing *in vitro* upon Treg activation (Figure 5.12). This may be explained by the reduced proliferation of Treg *in vitro* compared to Tresp [591, 592]. In addition, it has been suggested that CRNDE may have a role in pluripotency due to observations that CRNDE is highly expressed in ESCs [442] and spermatozoa [526] with its expression decreasing with differentiation [526]. The observation that CRNDE increases in expression during T cell differentiation counters this suggestion. However, considering the high cycling state of ESCs [593] and low cycling state of naïve T cells, these observations are in fact consistent and suggest a general role for CRNDE in cell cycling, proliferation and activation.

Despite the decrease in CRNDE expression in Treg upon activation and the requirement for cell activation for delivery of LNAs via gymnosis, I was able to achieve knockdown of CRNDE (Figure 5.10). In assessing the impact of loss of CRNDE on Treg phenotype, I saw a significant decrease in expression of the TNF receptor superfamily member CD30 (TNFRSF8) on Treg compared to a non-targeting control (Figure 5.11). However it should be noted that treatment with the non-targeting control lead to an increase in CD30 expression compared to treatment with no LNA suggesting that the difference in CD30 expression upon CRNDE knockdown may relate to an artefactual increase in expression in the non-targeting control. If CRNDE does truly contribute to CD30 expression on T cells, this may indicate a link between the role of CRNDE in proliferation and the role of CD30 in Treg expansion. The use of further control LNAs should help to resolve this issue.

Encouragingly, LOC286442 expression was associated with several genes relevant to Treg biology, supporting a role for the lincRNA in Tregs. These genes included *FOXP3, Helios* and *TNFR2*. However, I did not see a difference in expression of these proteins upon knockdown of LOC286442 using LNAs (Section 5.2.10). This may relate to the relatively low knockdown efficiency seen for LOC286442 compared to CRNDE (Figure 5.13). Additional experiments to probe the role of LOC286442 in Treg are suggested in Section 6.5 (Future work).

While knocking down *IFNG* super-enhancer eRNAs to assess their impact

on *IFNG* expression, I observed a change in expression of IFNγ by flow cytometry that correlated with the eRNA targeted (Figure 5.15). However this difference in IFNγ protein did not correlate with a difference at the transcript level and treatment with LNAs did not produce effective knockdown of the eRNA targets (Figure 5.16). This lack of knockdown is perhaps not unexpected for eRNA transcripts, which are unstable and low in abundance. In addition, as LNAs can sterically inhibit transcript activity as well as induce cleavage, it is possible that while a decrease in transcript level was not observed, eRNA function may nevertheless have been reduced.

6 Discussion

The lineage choice of a naïve CD4 T cell during differentiation supports the tailoring of the immune system towards different pathogens. Although the DNA sequence present in each T cell is nearly identical, the process of differentiation allows different T cell identities to arise through the increased activity of lineage-specific genes and the repression of genes relevant to alternative cell fates. These lineage-specific gene expression programs require maintenance at the level of transcriptional regulation. Transcriptional activity is governed by accessibility of DNA regulatory regions, such as enhancers, to regulatory proteins including transcription factors and chromatin-modifying complexes. This accessibility is in part influenced by histone modifications. Recently, the pervasive and tissue-specific transcription of multiple classes of ncRNAs has been demonstrated to play a role in cell differentiation and function and has altered our perception of the central dogma of molecular biology.

I have presented here the identification of several classes of ncRNAs with T cell lineage-specific expression and that have the potential for influencing T cell function. These include lincRNAs that are upregulated in primary human Tregs compared to Tresp and eRNAs emanating from T-bet super-enhancers that show Th1 specific expression. Through knockdown experiments, I have investigated the functional role of these ncRNAs. In addition, I have explored the dysregulation of Treg lincRNAs in the Treg relevant context of autoimmunity, specifically SLE. Differences in Treg gene expression between SLE patients and healthy individuals have also been explored, focused on transcription factors and changes in

histone modifications at genes dysregulated in SLE.

6.1 The role of CRNDE in Treg

The lincRNA CRNDE was identified as upregulated in Treg compared to Tresp (Section 3.2.2). Supporting a role for CRNDE in Treg, there is some enrichment of the Treg lineage-specifying transcription factor FOXP3 at the CRNDE gene suggesting its expression is regulated by this factor (Section 3.2.4). The gene TSS exhibits H3K4me3, while H3K4me1 is minimal (Figure 3.5), and contains introns, patterns that are consistent with *CRNDE* encoding a lincRNA rather than an eRNA. The expression of lncRNAs has been suggested to be a mechanism for the genomic targeting of chromatin-modifying complexes to specific genes (Section 1.3.3.1., [441]). Consistent with this, CRNDE has been shown to associate with the chromatin-modifying complexes PRC2 and coREST [441] and knockdown of PCR2 and CRNDE show overlap in the biological processes affected [441]. The repeats present in CRNDE exonic sequences (Figure 3.7) support the possibility of secondary structures that may promote an interaction with chromatin modifiers (Section 1.3.3). If CRNDE acts to regulate the expression of other genes, one might expect those genes to share a gene expression profile with CRNDE. The correlation between *CRNDE* expression and that of its genomic neighbour *IRX5* across several cell types **[526]** could therefore be indicative of regulation of *IRX5 in cis*. While this correlation is not seen here in Treg (Section 3.2.6) and IRX5 does not have a known function in T cell differentiation, other genes co-regulated with CRNDE are potential candidate target. With this hypothesis in mind, I examined the identities of genes that I found correlated with *CRNDE* expression (Section 5.2.1) and found several that had been

published to have roles in Treg function, including *CD30* and *4-1BB*. After optimising techniques to achieve knockdown of lncRNAs in T cells, I examined changes in the expression of these genes upon depletion of CRNDE (Section 5.2.9). I found a significant decrease in CD30 expression correlated with CRNDE knockdown, which implicates CRNDE in Treg expansion through supporting the expression of this TNF receptor that has a role in this process. However, due to the non-specific increase in CD30 seen with a non-targeting control LNA, it is possible that this is artefactual. Consistent with this, I did not observe any effect of CRNDE LNA on expression of the proliferation marker ki67. Experiments with additional negative control LNAs and additional independent LNAs targeting CRNDE will help resolve this issue.

Gene ontology analysis showed that genes with similar expression to CRNDE were enriched for processes relating to cell cycling and proliferation. This is consistent with the published association of CRNDE with several cancers [524, 548]. The upregulation of CRNDE *ex vivo* could therefore reflect the proliferation of Treg *in vivo* in response to self antigens [591, 592]. In SLE patients, although the difference was not significant, CRNDE showed a tendency towards decreased expression in memory compared to naïve Treg (Section 3.2.9). This may reflect a lower activity of memory Treg in SLE, which other published research has suggested (Section 1.1.3.6.1, [96]). The decreased expression of CRNDE in Tresp compared to Treg could be due to the increased mTOR signalling that is seen in Tresp [201, 594]. Ellis et al. 2014 observed that CRNDE expression decreased in fibroblasts upon insulin signalling through mTOR and Raf/MAPK [550]. In addition, *CRNDE* expression was shown to

increase through treatment with rapamycin, an inhibitor of mTOR, even in the absence of insulin signalling [550]. This is relevant to Treg biology because rapamycin promotes the expansion of Treg [201, 595] and therefore CRNDE may have a role in supporting Treg expansion. Ellis et al. 2014 showed this relationship between mTOR and CRNDE was unique to unspliced transcripts containing intronic sequences rather than spliced transcripts [550]. With this in mind, I performed isoform specific qPCR, demonstrating the presence of spliced isoforms of CRNDE that contributed to the difference in expression between Treg and Tresp (Section 0). This suggests that CRNDE may not be regulated downstream of rapamycin in Tregs. However, as control of the CRNDE gene may differ between fibroblasts and T cells, it remains to be seen if CRNDE expression is Treg is responsive to rapamycin.

6.2 The role of lincRNA LOC286442 in Treg

The lincRNA LOC286442 was also upregulated in Treg compared to Tresp and showed the highest fold difference of any ncRNA by array (Section 3.2.1). Similar to the *CRNDE* gene, FOXP3 bound to the promoter of *LOC286442* suggesting it forms part of the Treg transcriptional program. In addition, LOC286442 exhibited lower (but not significantly so) expression in Treg relative to Tresp in SLE patients compared to healthy individuals (Section 3.2.9), indicating either decreased expression of LOC286442 in SLE Treg and/or increased expression in SLE Tresp. Hierarchical clustering identified genes with similar expression patterns to *LOC286442* as enriched for processes related to T cell differentiation and lymphocyte development. This included the Treg genes *FOXP3, Helios* and *TNFR2*. Treatment with LNAs targeted against LOC286442 resulted in modest

knockdown of the lncRNA, but no effect on the expression of these proteins.

6.3 The functional importance of super-enhancer transcription

I have demonstrated here (Chapter 4) that T-bet super-enhancers exhibit lineage-specific transcription that correlates with transcription from associated Th1-specific genes. These super-enhancer eRNAs share characteristics with previously annotated functional enhancer transcripts such as lack of poly-adenylation [420, 421] and bidirectional transcription [420, 421]. In addition, I determined that eRNA transcription is increased at T-bet super-enhancers compared to normal enhancers and I detailed the chromosomal coordinates of eRNAs upregulated in Th1 cells from T-bet super-enhancers. At the *IFNG* super-enhancer, I demonstrated using a knockout mouse that loss of T-bet leads to reduced transcription of these eRNAs, in addition to suboptimal transcription of the *IFNG* gene itself. Additionally, observations from our lab have show that the transcriptional elongation factor P-TEFb is recruited to super-enhancers and this binding localises with that of T-bet. I have shown here through chemical inhibition that these eRNAs are a product of that P-TEFb activity. These data are consistent with lineage-specific control of gene expression by super-enhancers being mediated through recruitment of P-TEFb by T-bet with the resultant transcription of eRNAs.

Following these observations, I questioned whether the presence of the super-enhancer eRNA transcripts themselves are a functional requirement for transcription of downstream genes. Using the super-enhancer upstream of *IFNG* as a model for super-enhancer eRNA function, I

performed LNA knockdown experiments to assess the impact of the loss of these eRNA transcripts upon *IFNG* transcription (Section 5.2.11). The technical difficulties involved in knocking down eRNAs are appreciated within the field, as demonstrated by the low frequency of reports demonstrating functional roles for eRNAs. These difficulties arise from the nuclear localisation of eRNAs, which makes them relatively impervious to classical knockdown techniques, and their low abundance and low stability, which decrease the likelihood that knockdown methods will have any effect. In Th1 polarising cultures from two separate donors treated with 8 different LNAs against 4 eRNAs, I was unable to achieve knockdown that corresponded with a decrease in IFNG transcript or protein. Despite no detectable change in eRNA or IFNG transcript, the level of IFNγ protein did appear to differ in a manner that correlated with the targeting of LNAs, albeit with donor variation. As LNAs can limit transcript activity in addition to inducing RNaseH mediated cleavage, it is conceivable that LNA treatment may have affected the activity of these eRNA transcripts leading to changes in IFNG expression. It is also theoretically possible that a change in IFNG mRNA not detectable in the assays presented here may have resulted from alternative splicing of the *IFNG* gene. While regulation of alternative splicing has been functionally attributed to ncRNAs such as the *Zeb2* natural antisense transcript [448], to my knowledge such as role has not been previously postulated or supported for eRNAs or even upstream enhancers, although looping of the *IFNG* locus may provide a platform for interactions between super-enhancer eRNAs and splicing regulatory regions within the gene.

6.4 The transcriptional identity of SLE Treg

Through investigating differential gene expression in naïve and memory Treg of SLE patients compared to healthy individuals, I identified several transcription factors that were differentially expressed (Section 3.2.8). This included MAFB, which was upregulated in SLE Treg and has been reported in T cells before [545] but does not have a well understood role, and NR4A3, which has recently been demonstrated to support Treg development [547]. Neither of these transcription factors have been associated with SLE or any other human autoimmune disease before. From the work presented here it is not possible to determine whether the increase in the expression of these transcription factors could be responsible for a Treg defect that may contribute to pathogenesis or whether their increased expression instead reflects the activity of Treg in responding to inflammation.

I also found that H3K9me3 was significantly increased in Treg and Tresp at genes that were differentially expressed between SLE patients and healthy individuals (Section 3.2.10). This was unexpected because the majority of literature relates H3K9me3 to chromatin condensation for the repression of genomic instability at repetitive genomic regions [361, 362, 364]. An additional role for H3K9me3 has been described at actively transcribed genes [363] through regulation of alternative splicing [371]. This may suggest the differential expression seen in SLE may be in part regulated at this level and supports further investigation of the control of H3K9me3 deposition and its function in cell lineage specification in response to environmental stimulus.

6.5 Future work

I have presented here several knockdown approaches to assess the impact of loss of ncRNA on cell function (Chapter 5). LNA gapmer oligonucleotides, which exert their effect through RNaseH-mediated cleavage of DNA-RNA hybrids, were successful in knocking down lincRNAs that were identified as preferentially expressed in Treg (Chapter 0). However, LNAs did not successfully knockdown eRNAs upstream of *IFNG* (Chapter 4). As LNAs can also interfere with target activity and function without cleavage of the transcript, it remains possible that LNAs could be used to inhibit eRNAs. Observing this effect is challenging compared to detecting changes in transcript level by qPCR but I did not detect an effect of LNAs on *IFNG* expression suggesting the LNAs did not affect eRNA function. However this does not rule out a functional relationship between *IFNG* super-enhancer eRNA transcripts and expression of *IFNG* and this could be explored further through tiling of LNAs along the length of all *IFNG* super-enhancer eRNAs in order to target all potentially functional transcript regions. In addition, these eRNAs may be dynamically regulated throughout Th1 differentiation and so characterisation of their transcription over a time may aid in defining the point in differentiation when functional effects are likely to be seen. In addition, recent evidence has demonstrated the presence of Dicer in the nucleus [596, 597], at least in some mammalian cellular contexts, suggesting an shRNA approach may be more effective than previously thought. However, attempts have been made in our lab to knockdown these eRNAs using shRNAs without success (Arnulf Hertweck). Genome editing using CRISPR/Cas (clustered regularly interspaced short

palindromic repeats/CRISPR associated) or TALEN (transcription activator-like effector nucleases) systems may provide an alternative, more robust strategy. CRISPR/Cas is based on the disruption of genomic sequence through the induction of double strand breaks catalysed by the Cas9 endonuclease, with CRISPR RNA sequence guiding the site of this cleavage [598, 599]. TALEN are similar in principle, with nuclease activity being guided by the specific binding of a fused TAL effector DNA binding domain [600]. Applied to protein-coding genes, genome editing allows targeted disruption of open reading frames and substitution of codons to render a protein dysfunctional. This technique can also be aimed at disruption of transcriptional start sites in order to ablate transcription entirely, which is relevant for non-coding regions that produce transcripts in which functional domains are unknown.

It is possible that the identity and sequence of transcripts produced from the *IFNG* super-enhancer has less consequence on expression of *IFNG* than the act of transcription itself. For example, eRNA transcription may allow binding of other factors to enhancers or allow histone modifications important for enhancer activity. A potential mechanism would be through the formation of RNA:DNA hybrids termed R-loops [601]. R-loops are triplex structures involving two strands of single stranded DNA unwound as a result of RNAP activity with a single strand of complementary nascent RNA annealed to one strand [601]. Among other transcriptional processes, these have been implicated in the recruitment of histone methyl transferases leading to tri-methylation of H3K4 upstream of transcription start sites [602, 603]. Detection of R-loops at transcribed super-enhancer

sites could be performed by nucleic acid isolation coupled to RNaseH mediated cleavage and sequencing or using the R-loop-specific antibody S9.6 [602].

I have demonstrated here that production of eRNA from T-bet super-enhancers co-localises with T-bet and P-TEFb binding (Chapter 4, Figure 4.8). In addition, increased transcription is a function of the increased T-bet binding at super-enhancers compared to typical enhancers (Figure 4.10) and this transcription is reliant upon both T-bet (Figure 4.13) and P-TEFb (Figure 4.14Figure 4.15). In addition to potential functional roles for the transcripts produced from super-enhancers, this transcription may be a non-functional biochemical by-product of the binding of the transcriptional elongation factor P-TEFb. We have suggested that P-TEFb recruitment by T-bet is a mechanism for how T-bet and other lineage-specifying transcription factors promote differentiation (publication is currently in peer review process). Further support for this could come from demonstrating loss of P-TEFb binding at super-enhancers leads to reduced P-TEFb binding at downstream genes. This would require site specific inhibition of P-TEFb binding at super-enhancers but not at genes. This could be achieved through site specific disruption of P-TEFb bound regions at super-enhancers through genome editing techniques. In addition, RNA-Seq from cells treated with P-TEFb inhibitors (JQ1 and flavopiridol, Section 4.2.8) would indicate if P-TEFb regulated transcription is specifically disrupted at lineage specific genes. As JQ1 and flavopiridol are already in clinical use to inhibit transcription in the context of several cancers [604, 605] their application to inflammatory diseases could also be realised. As

super-enhancers are enriched for disease associated SNPs [326], it would also be interesting to determine if Th1 eRNAs are associated with SNPs linked to autoimmune and allergic conditions.

I have identified several long ncRNAs that are upregulated in primary human Treg compared to Tresp (Chapter 3). CRNDE and LOC286442 were investigated in functional assays through knockdown with LNA gapmer oligonucleotides (Chapter 5). While this approach achieved successful knockdown, phenotypic effects, specifically the expression of biologically relevant genes with similar expression patterns, were not observed. Additional approaches could focus on identifying perturbations in expression of cytokine and chemokine receptors and the secretion of cytokines and chemokines, particularly those with similar expression to CRNDE (such as IL7, CCRL2, IL12RB2, IL1R1, CCR10, CCR3, CCR4, CCR5 Table 7.31) and LOC286442 (such as IL21, IL1RAP Table 7.32). In addition, examination of the effect of knockdown on global gene expression, rather than the identification of genes with similar expression to the target of interest *ex vivo*, could be achieved using microarrays. Ultimately, the effect of CRNDE expression of Treg function could be demonstrated through *in vitro* Treg suppression assays, measuring the effect of Treg on Tresp proliferation and cytokine production. The effect of knockdown on differentiation rather than maintenance of Treg could also be explored, for example through *in vitro* polarisation or expansion of Treg. The latter is particularly relevant as this is often performed with rapamycin [201, 595], which has been shown to increase the level of CRNDE transcripts in fibroblasts [550]. In this context, rapamycin was used as an inhibitor of insulin signalling through blockade of mTOR, which decreased

expression of unspliced transcripts of CRNDE. While I have seen preferential expression of spliced rather than unspliced transcripts in Treg(Section 0), the influence of rapamycin on CRNDE may still have significance in this alternative cellular context. It would therefore be informative to explore the effect that rapamycin has on CRNDE expression and the effect of CRNDE knockdown in Treg or overexpression in Tresp during expansion conditions.

CRNDE expression *ex vivo* was shown to correlate with biological processes of cell cycling and proliferation (Section 5.2.1). However, expression of CRNDE decreased rather than increased in Treg with *in vitro* activation, which may be related to the decreased proliferation of Treg observed *in vitro* [591, 592]. This complicates interpretation of LNA-mediated knockdown experiments, which require cell activation for uptake of LNA. To circumvent this issue, the use of a recently published C*rnde* knockout mouse [410] could be used to assess loss of Crnde *in vivo*. Induction of systemic inflammation, for example with Complete Freund's Adjuvant (CFA), could allow investigation of the role of Crnde in immune activation. In addition, the effect of CRNDE overexpression in Tresp could be explored but this has limitations if the function of CRNDE is dependent upon the genomic site of its transcription.

The cellular localisation of ncRNAs is directly related to their mechanisms of action. For example, those with roles in transcriptional regulation of gene expression are located in the nucleus (such as Xist [398]) while cytoplasmic ncRNAs may have roles in structure and protein localisation (such as NRON [462]). Therefore, confirming the nuclear localisation of CRNDE and LOC286442, through subcellular fractionation and qPCR,

would support a role for the impact of these ncRNAs on transcriptional regulation of gene expression in Treg. Structural predictions would also be informative of ncRNA mechanism, as would experiments identifying proteins with which the RNAs associate.

By comparison of expression between naïve and memory Treg and Tresp of SLE patients and healthy controls, the role of the identified Treg ncRNAs in the context of autoimmunity was also explored (Section 3.2.9). Although none were significantly up or down regulated in SLE patients, CRNDE showed the greatest median $\log_2$ fold difference in naïve compared to memory Treg between SLE patients and healthy individuals (median $\log_2$ ratios of 0.6 and -1.6 respectively, Figure 3.13.B) suggesting it is either upregulated in naïve Treg and/or downregulated in memory Treg in SLE. Given the association of CRNDE with cell cycling and proliferation (Section 5.2.1) and observations that memory Treg ($FOXP3^{high}CD45RA-$, 'activated Treg' [96]), which are more suppressive [96], are present at a lower frequency in SLE patients (Section, 1.1.3.6, [96]), it may be informative to correlate the expression of CRNDE in memory Treg with disease activity SLE patients. This may be achievable in a larger cohort of SLE patients through *in situ* hybridisation coupled to flow cytometry.

Several differentially expressed transcription factors in Treg between SLE patients and healthy controls were identified (Figure 3.12). Of these, NR4A3 (nuclear receptor subfamily 4, group A, member 3, also known as NOR1) has been functionally associated with Treg, with the NR4A family essential to Treg development [547]. Investigation of the role of NR4A3 upregulation in SLE would require determination of whether this differential expression precedes autoimmune inflammation or is a consequence of it.

This could be achieved through comparison of NR4A3 expression in Treg of SLE patients and healthy controls in response to *in vitro* culture with inflammatory cytokines.

While identification of differential histone modification patterns between SLE patients and healthy individuals was precluded by the low frequency of Treg in peripheral blood and the relatively high cell numbers required for ChIP-Seq, differential histone modifications were identified at genes that differed in expression between SLE patients and healthy individuals (Section 3.2.10). Enrichment of H3K9me3 was seen at genes upregulated in healthy individuals compared to SLE patients (Figure 3.15 and Figure 3.16). As H3K9me3 has been associated with regulation of splicing during transcriptional elongation [363, 371], it would be informative to see if these genes are alternatively spliced between SLE patients and healthy individuals. This may also suggest that regulation of splicing through H3K9me3 deposition is influential in determining a response to an inflammatory environment. This could be further investigated by observing enrichment of H3K9me3 genome wide in response to *in vitro* culture with inflammatory cytokines. Recent developments allowing ChIP-Seq with tens of thousands of cells [606] also now allow for histone modifications to be measured in Tregs from healthy versus SLE patients.

7 Appendix

7.1 Differentially expressed transcripts between Treg and Tresp identified by microarray

7.1.1 Differential gene expression between T cell subsets in healthy donors

Table 7.1: Transcripts upregulated in total Treg compared to total Tresp

Gene symbol			
A_33_P3298398	E2F7	IL2RB	PLAU
A_33_P3332616	ENST00000354689	IL7	PLCG2
ABHD5	ENST00000390247	INPP5F	PLEKHG1
ACTA2	ENST00000390252	INTS6	PMAIP1
ADAMTS4	ENST00000390559	IRAK3	POU2AF1
ADRBK2	ENST00000390606	KALRN	PTPLA
AHR	ENST00000432978	KCNK1	PTTG1
AIM2	ENST00000444919	KIAA0101	PTTG2
AP3M1	ENST00000477036	KIF20B	PTTG3P
AR	ENST00000492167	KIF5B	RANBP9
ARG1	ENST00000502284	KLK1	RBBP8
ARHGAP11B	ENST00000517927	LAIR2	RBMS3
ARHGAP31	ENTPD1	LAT2	REL
ASPM	F5	LAYN	RPS16P5
ATP2B1	FAM102B	lincRNA:chr10:118574085-118608460_R	RRM2B
ATP2B4	FAM129C	lincRNA:chr10:97639610-97667003_F	RTKN2
ATXN7L1	FAM160B1	lincRNA:chr12:12884108-12894958_F	SELP
BARD1	FAM19A2	lincRNA:chr12:32571608-32576908_R	SEMA7A
BASP1	FAM46C	lincRNA:chr13:50614546-50615364_R	SESTD1
BCAS1	FANCA	lincRNA:chr18:68297754-68318093_R	SETD7
BIRC3	FANCL	lincRNA:chr2:179914333-179915343_R	SGPP1
BIRC5	FANK1	lincRNA:chr2:213437296-213768199_R	SHMT2
BRCA2	FBXL20	lincRNA:chr2:213703255-213723530_F	SKAP2
BRIP1	FCRL1	lincRNA:chr2:64455559-64478475_F	SLC14A1
BUB1	FCRL2	lincRNA:chr21:30749379-30901454_F	SLC2A8
C18orf1	FCRL3	lincRNA:chr3:106526585-106630535_R	SLC39A10
C18orf54	FHL3	lincRNA:chr3:106619835-106630535_R	SLC9A7
C9orf167	FLJ45482	lincRNA:chr3:114031960-114042235_F	SLC9A7P1
CASC5	FOXP3	lincRNA:chr3:13093175-13147525_R	SMC4
CASK	GADD45A	lincRNA:chr7:22611226-22701411_R	SMC6
CCDC50	GCC2	lincRNA:chr7:7928975-7996750_R	SMPD3
CCNA2	GCNT4	lincRNA:chr8:95917724-95933974_F	SNORA16B
CCNB2	GEN1	lincRNA:chrX:115401722-115413847_R	SPATS2L
CCNG2	GJB6	LMCD1	STAM
CCR3	GK	LOC100131234	SUOX
CD200	GLCCI1	LOC100132741	SWAP70

Gene symbol			
CD22	GNB4	LOC286442	TBC1D4
CD24	GPD2	LRRC32	TBC1D9
CD38	GPR174	LRRC37A3	TFRC
CD40	GPR55	LUZP6	TIGIT
CD74	GSTA4	MAP3K5	TK1
CDCA5	HECW2	MCM4	TMEM38A
CDCA7	HELLS	MELK	TNFRSF13C
CDCA8	HIST1H1B	METTL7A	TNFRSF1B
CDHR3	HIST1H2AI	MGST2	TNFRSF8
CDK1	HIST2H3A	MINK1	TNFRSF9
CDK14	HJURP	MKI67	TOX
CEACAM4	HLA-DMA	MREG	TP53INP1
CENPE	HLA-DMB	MXD3	TPR
CENPI	HLA-DOA	MYB	TRIB1
CENPM	HLA-DPA1	MYO1E	TRIM32
CEP128	HLA-DPB1	MYO5C	TRIO
CEP55	HLA-DQA1	NAMPT	TTN
CHDH	HLA-DQA2	ND4L	TUFT1
CHST7	HLA-DQB1	NEDD4L	TXNDC11
CIITA	HLA-DQB2	NF1	TYMS
CIT	HLA-DRA	NHLRC2	UBE2C
CLNK	HLA-DRB1	NP113779	UHRF1
CNGA1	HLA-DRB4	NUSAP1	USP15
CNKSR2	HLA-DRB5	OAS1	UTS2
CRNDE	HMGB3	OBFC2A	VAV3
CTLA4	HMMR	OGN	VDR
CTNNAL1	HS3ST1	OIP5	WNT10A
CTPS2	HVCN1	P2RY1	XM_003118559
CTSC	ICA1	PARVB	ZBTB32
CXXC5	IDE	PBK	ZC3H12D
DNAH8	IKZF2	PDE4D	ZNF193
DOK3	IKZF3	PELI1	ZNF532
DTL	IKZF4	PER2	ZNRF2
DUSP10	IL1R2	PICALM	ZNRF2P1
DUSP4	IL2RA	PIP5K1B	

Table 7.2: Genes downregulated in total Treg compared to total Tresp

Gene symbol			
A_33_P3253902	CTSW	lincRNA:chr3:184997356-185017181_F	RSAD2
A_33_P3271187	CYB561	lincRNA:chr3:184997356-185033531_F	RTN4R
A_33_P3343605	CYP2J2	lincRNA:chr3:18507452-18568034_F	SBF2
ABCB1	D4S234E	lincRNA:chr6:11118364-11146189_R	SCARNA6
ABLIM2	DENND5A	lincRNA:chr7:130793407-130794935_R	SCML4
ACR	DHRS3	LMF1	SDC3
ACSL6	DOCK9	LOC100130000	SEMA4C

Gene symbol			
ACVR1C	DPP4	LOC100130298	SFMBT2
ADAM23	DUSP14	LOC100506342	SH3BGR
ADC	EDA	LOC100507199	SLC40A1
ADCY9	EFHC2	LOC440297	SMAGP
ALPK1	EMR4P	LOC441461	SNAPC1
ANK3	ENST00000390341	LPIN2	SOAT2
ANKRD55	ENST00000414198	LRRC6	SOCS3
APBA2	ENST00000449914	LRRN3	SOX8
ARHGEF4	ENST00000512129	MATK	SPEF2
ATP10A	EOMES	MDGA1	SPEG
AXIN2	FAM19A1	ME3	SPEM1
B3GNT3	FCGBP	MORC2-AS1	ST6GALNAC1
BAG3	FGF9	MSRA	SUN3
BC036215	FRY	NAP1L3	SVIL
BEX5	GCET2	NELL2	SYTL2
C12orf53	GPRC5B	NFIA	TAF4B
C14orf132	GRASP	NLRP3	TC2N
C14orf45	GRB10	NMUR2	TCEAL2
C17orf67	ID2	NOSIP	THC2564554
C1orf151-NBL1	IL4R	OTUD7A	THC2704282
C1orf172	KCTD13	PCSK5	THEMIS
C2orf40	KCTD3	PDE4D	TLR1
CACNA1B	KRT2	PDE6B	TLR2
CAND2	KRTAP4-11	PIM1	TMEM63C
CASP10	lincRNA:chr1:173831852-173837202_F	PION	TMEM71
CCDC112	lincRNA:chr1:59049237-59065662_R	PITPNC1	TRPV5
CCL27	lincRNA:chr10:4944575-4977950_F	PLAG1	TTC16
CD200R1	lincRNA:chr10:89908970-89919320_F	PLCB1	UGT3A1
CDK6	lincRNA:chr11:8697549-8703349_F	PLEK2	VIPR1
CEBPE	lincRNA:chr12:102593870-102602595_F	PLXDC1	WDR86
CELA1	lincRNA:chr12:127160147-127258072_F	PRKCE	WNT1
CFH	lincRNA:chr14:71020897-71044497_F	PRRT2	WNT10B
CFHR3	lincRNA:chr14:92220122-92237072_F	PTGDR	WNT7A
CHD7	lincRNA:chr14:92223102-92223962_R	PTPRO	WNT7B
CHN2	lincRNA:chr14:96507172-96661947_F	RAD54B	XYLT1
CHRNB3	lincRNA:chr14:96548697-96561747_F	RBMS1	ZACN
CNIH4	lincRNA:chr14:96555696-96556712_F	RECK	ZBTB47
CNTD1	lincRNA:chr15:40828033-40834483_F	RIPK2	ZNF462
COPS7B	lincRNA:chr16:90043749-90053849_R	RLN2	ZNF541
CR1	lincRNA:chr2:105424027-105429249_F	ROBO3	
CR2	lincRNA:chr20:56156369-56171644_F	RORB	
CRB3	lincRNA:chr22:25588075-25594350_F	RRAGD	

Table 7.3: Gene upregulated in healthy Treg naive compared to Treg memory, Tresp naïve and Tresp memory

Gene symbol			
AF194718	ENST00000390237	FCRL2	MS4A1
AFF3	ENST00000390243	FCRL5	NCF1
BANK1	ENST00000390294	FCRLA	NP113779
BCL11A	ENST00000390551	FXYD2	OSBPL10
BF175071	ENST00000390556	HHEX	PLCG2
BLNK	ENST00000390603	HS3ST1	PNOC
CD19	ENST00000390615	HVCN1	SETBP1
CD22	ENST00000390632	KIAA0125	SNX22
CD24	ENST00000424969	KIAA0226L	SPIB
CDCP1	ENST00000426099	LILRB1	SYK
CNR2	ENST00000468879	lincRNA:chr11:117668490-117680359_R	TBC1D9
COBLL1	ENST00000471857	lincRNA:chr2:165531204-165538379_R	TCL1A
CPNE5	ENST00000479981	LOC283663	TLR10
CXXC5	ENST00000492167	LOC96610	TNFRSF13C
EBF1	FAM129C	LYN	VPREB3
ENST00000354689	FCRL1	MEF2C	WDFY4

Table 7.4: Genes downregulated in healthy Treg naive compared to Treg memory, Tresp naïve and Tresp memory

Gene symbol
Zero transcripts identified

Table 7.5: Genes upregulated in healthy Treg memory compared to Treg naïve, Tresp naïve and Tresp memory

Gene symbol			
ACTA2	E2F2	lincRNA:chr2:111965352-112252622_F	RHOXF1
ADPRH	ENTPD1	lincRNA:chr2:111965652-112101793_R	RTKN2
ANKS1B	EZH2	lincRNA:chr2:111992304-111994467_R	SCG5
BARD1	F5	lincRNA:chr2:112005726-112123291_R	SEMA3G
BC021693	FAM110C	lincRNA:chr3:171506056-171528740_R	SGMS1
BFSP2	FAM124B	lincRNA:chr3:8615882-8616354_F	SLC1A1
BIRC5	FAM160B1	lincRNA:chr7:105551484-105564551_F	ST7
C15orf53	FAM164A	lincRNA:chr7:105551628-105553084_F	STAM
CCR10	FAM46C	lincRNA:chrX:115401722-115413847_R	TFRC
CCR3	FANK1	LMCD1	THC2477815
CCR4	FAS	LMNA	TK1
CCR8	GINS2	LOC541471	TNFRSF8
CD70	GPR19	LRRC32	TNFRSF9
CDHR3	HPGD	MEOX1	TOX
CEP128	IL1R1	MYO5C	TRIB1
CHEK1	IL7	NEDD4L	TRIM16L
CHST7	INPP1	NINJ2	TTN
CLNK	JAKMIP1	OAS1	UTS2

Gene symbol			
CPNE2	JSRP1	OBFC2A	VANGL1
CTNNAL1	KIAA1841	PHTF1	VDR
CXCR6	LAYN	PMAIP1	WEE1
DGKH	LINC00152	PMCH	ZBTB32
DUSP10	LINC00312	PMCHL1	ZC3H12C
DUSP4	lincRNA:chr2:111954496-112125102_R	PTPLA	

Table 7.6: Gene downregulated in healthy Treg memory compared to Treg naïve, Tresp naïve and Tresp memory

Gene symbol			
ACVR1C	DACT1	lincRNA:chr8:32785489-32902343_F	PLAG1
ALPK1	DCHS1	LMF1	PLXDC1
APBA2	DENND5A	ME3	RASGRF2
AXIN2	FAM26F	NAP1L3	SATB1
C2orf40	FBP1	NELL2	SLC40A1
CERS6	IGF1R	NRIP1	SPG20
CRTAM	KCTD3	PCSK5	TCF7
CYP2J2	lincRNA:chr10:4944575-4977950_F	PELI2	THC2564554

Table 7.7: Genes upregulated in healthy Tresp naive compared to Treg naïve, Treg memory and Tresp memory

Gene symbol	
DACT1	KRT2

Table 7.8: Genes downregulated in healthy Tresp naive compared to Treg naïve, Treg memory and Tresp memory

Gene symbol			
ACTA2	FANCL	lincRNA:chr10:118574085-118608460_R	PYHIN1
ADAM19	FAS	lincRNA:chr21:26934779-26952204_F	SETD7
AHNAK	FCRL3	lincRNA:chrX:115401722-115413847_R	SGPP2
AIM2	ICA1	LOC338620	ST8SIA1
ATXN1	IL18R1	MAP3K5	TIGIT
BACE1	IL2RA	NCF4	TNFRSF1B
CHST7	IL2RB	PERP	TP53INP1
FAM46C	ITGB1	PREX1	WNT10A
FANCA	LAIR2	PTPN22	ZNF193

Table 7.9: Genes upregulated in healthy Tresp memory compared to Treg naïve, Treg memory and Tresp naïve

Gene symbol			
ABCB1	CFH	IL15	PLCB1
ADAM23	CFHR3	IL18RAP	PRR5L
ADRB2	CR1	IL4I1	PTGDR
AUTS2	D4S234E	LATS2	SOAT2

Gene symbol			
BFSP1	F2R	LYAR	SPON1
BHLHE40	HOPX	MCOLN2	STOM
CCL5	IFNG	MYBL1	TMEM200A

Table 7.10: Genes downregulated in healthy Tresp memory compared to Treg naïve, Treg memory and Tresp naïve

Gene symbol		
CD38	lincRNA:chr2:179914333-179915343_R	PTGIR
GNG8	MYB	TTN

7.1.2 Differential gene expression between SLE patients and healthy controls

Table 7.11: Genes upregulated in SLE total CD4

Gene symbol			
A_24_P49597	FAM22G	lincRNA:chr9:37579075-37586275_R	RBM38
A_24_P67552	FAM46A	lincRNA:chrX:71132475-71339425_R	RBM44
A_33_P3229527	FANCC	lincRNA:chrX:73164162-73291277_F	RCL1
A_33_P3309643	FBXO32	lincRNA:chrX:73794800-73800175_R	REC8
A_33_P3402783	FBXO33	LOC100131726	REL
ABCA1	FCGR1B	LOC100506036	RELL1
ABCD3	FEM1B	LOC100509175	RIPK2
ACSS1	FLJ22447	LOC254057	RNF103
ADNP2	FLJ45684	LOC283575	RNMT
AHNAK	FOSL2	LOC374890	RPS26
AK074144	FTSJD2	LOC388796	RSAD2
AKR7L	FUT10	LOC401022	S100A12
ANK3	G0S2	LOC643770	S100A9
APOBEC3A	GADD45A	LOC649294	SCARF2
APOBEC3B	GFRA1	LOC650293	SEC14L3
APOM	GHRL	LOC729867	SECISBP2L
AQP9	GMFB	LOC79015	SEPSECS
AREG	GRASP	LPP	SETD7
ARID5A	GRM2	LRRC24	SH3YL1
ARL8B	GTF3C4	LRRC25	SIK1
ARRDC4	H1FX	LRRC8B	SLC16A6
ARTN	H3F3B	LY6E	SLC25A13
AVPI1	HCG4B	LYN	SLC25A30
B3GNT2	HERC5	MAFB	SLC2A3
B7H6	HERC6	MAFF	SLC3A1
BAALC	HIF1A	MAP3K8	SMAD1
BACH1	HIVEP3	MAX	SMURF1
BCL2	HMGB2	MBIP	SOD2

Gene symbol			
BCL2L11	HSH2D	MBP	SPATA24
BCL3	IDS	MCM9	SPATS2L
BCL6	IFI27	MED23	SPPL3
BE613848	IFI44	METRNL	STAT3
BGN	IFI44L	MGP	SUB1
BZW1	IFI6	MMD	SUGP1
C10orf68	IFIT3	MMP9	SUPV3L1
C10orf76	IFIT5	MRPL19	SYCE2
C1S	IFITM3	MRPL42P5	SYF2
C20orf111	IGFALS	MTFP1	SYNM
C20orf7	IKZF5	MTRR	TAB3
C5AR1	IL17C	MX1	TAF8
C7orf40	IL1RN	MX2	TAS2R20
C8G	IL20RB	MXI1	TBC1D15
CBX3P2	ILKAP	NAMPT	TEPP
CCDC157	IRS2	NAP1L5	TET2
CCDC19	ISG15	NCAPH	THC2489251
CCL2	ITCH	NEDD9	THC2499144
CCR1	ITGA7	NFIL3	THC2524986
CCT6B	ITGB2	NFKBIA	THC2566834
CD83	KCNK7	NFKBIB	THC2596076
CDC23	KLF10	NFKBIE	THNSL2
CDKN2D	KLF11	NFU1	THUMPD2
CDYL2	KLHDC7B	NPDC1	TMEM105
CEP95	KLHL12	NR4A3	TMEM119
CFD	KLHL15	NR6A1	TMEM2
CFL2	KLHL5	NRCAM	TMEM87A
CIR1	LAMP3	NUDCD3	TNFAIP3
CLEC7A	LAP3	NUDT4	TNNC2
CLP1	LATS1	OAS1	TOX
CMPK2	LINC00487	OAS3	TREM1
CNP	lincRNA:chr1:207534352-207540427_F	OCLN	TRIM22
COQ10B	lincRNA:chr10:119344760-119356235_R	ODC1	TSTD2
CPLX3	lincRNA:chr10:128586385-128592960_F	ODF3B	TXNDC11
CR605633	lincRNA:chr10:3521025-3527500_R	OR52N4	UBAP2L
CSDA	lincRNA:chr12:92366844-92524394_R	OXR1	UPB1
CTBS	lincRNA:chr12:92378756-92535489_R	PCGF5	USP18
CXCL2	lincRNA:chr13:48450924-48481320_R	PDE4D	USP41
CYB5RL	lincRNA:chr14:103209147-103227872_R	PELI1	UTP23
DCTN6	lincRNA:chr16:10479917-10481598_F	PER1	UTRN
DDIT4	lincRNA:chr17:78301614-78306077_F	PFKFB3	VAV3
DDX60	lincRNA:chr19:57769688-57777413_F	PGAP1	WASH3P
DEF8	lincRNA:chr19:57773281-57774106_F	PGGT1B	WHAMM
DENND3	lincRNA:chr2:216583792-216584130_R	PHLDB3	WIPF1
DENR	lincRNA:chr2:238021286-238040461_F	PIAS2	WTAP

Gene symbol			
DHX40	lincRNA:chr2:54284416-54307601_R	PIGA	XAF1
DUSP5	lincRNA:chr2:61157504-61158747_F	PIM3	XLOC_001688
DYNLT1	lincRNA:chr2:91693289-91693947_R	PLA2G4C	XLOC_006925
EDN1	lincRNA:chr3:128154578-128156195_F	PLBD1	XM_003118559
EFCAB10	lincRNA:chr3:188598106-188603231_F	PLK3	XM_003118785
EIF2AK2	lincRNA:chr3:194190336-194206786_R	PLSCR1	XM_003119830
ENST00000369381	lincRNA:chr3:47207196-47217796_F	POLR1C	YAF2
ENST00000370708	lincRNA:chr4:5529699-5536524_R	PPM1A	YOD1
ENST00000418620	lincRNA:chr5:131804581-131808735_F	PPM1B	ZBP1
ENST00000422603	lincRNA:chr5:139117291-139127116_R	PPT1	ZBTB43
ENST00000433105	lincRNA:chr5:179091169-179097344_R	PRB1	ZBTB44
EPSTI1	lincRNA:chr6:30001621-30013571_R	PRB4	ZC3H12A
ERC1	lincRNA:chr6:56749266-56765891_F	PRH2	ZFP36L2
ERMN	lincRNA:chr6:64472483-64499031_F	PRPH	ZNF292
ETNK1	lincRNA:chr7:12762625-12862575_R	PSMD5	ZNF295
ETV3	lincRNA:chr7:36771855-36772721_R	PTP4A1	ZNF326
EXOSC7	lincRNA:chr8:118761334-118761831_R	RAB11FIP1	ZNF507
EXT1	lincRNA:chr8:118762879-118763140_R	RAB30	ZNF584
FAM132A	lincRNA:chr8:126718218-126979293_F	RAD1	ZNFX1
FAM133B	lincRNA:chr8:6554042-6564617_R	RAD23B	ZRSR2
FAM13A	lincRNA:chr9:34666200-34672975_R	RAMP1	
FAM183B	lincRNA:chr9:36456325-36463750_R	RBBP6	

Table 7.12: Genes downregulated in SLE total CD4

Gene symbol			
A_33_P3357626	EPHB4	LOC100129186	SLC23A3
A_33_P3363305	ERAP1	LOC100131094	SLC24A1
A_33_P3382887	ETNK1	LOC100131138	SLC25A19
A_33_P3398196	F2RL1	LOC100132057	SLC35A1
A_33_P3399755	FAM118B	LOC100506676	SLC35B3
A_33_P3422712	FAM173B	LOC100506713	SLC35C1
AA489744	FAM203A	LOC151174	SLC35F3
ABHD6	FAM43A	LOC157562	SLC46A1
ACADM	FAM63A	LOC253039	SLC8A3
ACCS	FANCE	LOC283856	SPATA21
ACN9	FASLG	LOC389834	SPIN4
ACOT13	FASTKD1	LOC401397	SPTBN1
ADORA2A	FBLN5	LOC401431	ST6GAL1
AHSA1	FBXL18	LOC642852	STT3A
AK123308	FBXO44	LOC643837	SUCLA2
AK123945	FCRLB	LOC645431	TAF1A
AK124695	FGF9	LOC654433	TBC1D20
AKTIP	FIGNL1	LOC727916	TCEAL1
ANKH	FIGNL2	LOC728705	TDP1
ANKRD13C	FLCN	LOC728875	TDRD10

Gene symbol			
ANKRD5	FLJ33630	LOC729678	TEKT4
ANKRD58	FLJ39639	LPAR6	TERF1
AP1AR	FPGT	LRIF1	TES
ARFIP1	FSD1	LRR1	TFCP2
ARHGAP18	FTSJD1	LRRC27	THAP2
ARHGAP42	GBE1	LRRD1	THAP3
ARSK	GDF10	LY9	THAP7-AS1
ASPHD2	GDPD3	M6PR	THAP8
ATG4C	GEMIN6	MAB21L2	THC2537043
ATG9B	GFOD2	MAGOHB	THC2550350
ATP6V1C1	GGCT	MAT2A	THC2672097
ATP6V1D	GGPS1	MBOAT7	THC2724742
ATP9A	GIN1	MCMBP	THNSL1
ATXN10	GINS3	MED11	TIA1
BBS10	GKAP1	METTL12	TIMD4
BBS12	GLRX	METTL20	TKTL1
BC039426	GOLPH3L	MGAT2	TMEM14E
BPNT1	GOSR2	MGC2752	TMEM185A
BPTF	GTF2H1	MKKS	TMEM187
BREA2	GTF2H5	MKS1	TMEM218
BRK1	GVINP1	MMACHC	TMEM60
BTN2A1	HCG4	MMRN1	TMEM63A
BUD13	HIST2H2BE	MOB3A	TMEM79
C10orf57	HIST2H2BF	MON1A	TMEM8C
C12orf76	HKDC1	MRM1	TMEM99
C14orf109	HLTF	MRPS14	TNFRSF1A
C16orf5	HMGCR	MRS2	TOB2
C17orf69	HN1L	MSH2	TOR1A
C17orf80	HOMEZ	MSRB2	TOR1B
C18orf18	HOXB2	MTHFSD	TP53RK
C19orf25	HSCB	MTMR10	TPM1
C19orf73	HSD17B4	MYCBP	TRAF5
C1orf201	HSD17B7	MYOM2	TRIM13
C1orf31	HSPH1	N6AMT1	TRIM4
C1orf56	HYLS1	N6AMT2	TRIM68
C20orf134	ICMT	NAA30	TRMT1L
C2CD2L	IDI2-AS1	NAB2	TRNAU1AP
C2orf42	IFFO2	NAPEPLD	TTC36
C2orf43	IFNAR1	NDUFAF1	TTC9C
C2orf44	IHH	NEFM	TUBBP5
C5orf39	IKBIP	NEK4	TUBD1
C5orf62	ILK	NEURL	TUBE1
C6orf115	IPO5	NF1	TUBGCP4
C6orf211	IRAK2	NFYA	TXLNA
C8orf55	ITGA10	NMT1	TXNDC17

Gene symbol			
C9orf64	ITGB1BP1	NOC3L	TXNDC9
C9orf69	ITPR2	NPSR1	UBE2T
CASP10	JRKL	NS3BP	UBTD2
CASP8AP2	KBTBD6	NT5E	UFSP1
CBX1	KCNK6	NUP43	UGDH
CCDC121	KIAA1009	OPA3	UNG
CCDC126	KLHL25	OR7E5P	UTP14C
CCDC142	KLHL6	ORAI3	VEZT
CCDC34	KLHL9	ORC2	VPS35
CCDC75	KRCC1	ORMDL2	VSIG1
CCDC76	KRTAP1-3	OSBPL11	WDR41
CD200R1	KRTAP19-2	PAQR8	WDR62
CDK20	LACTB2	PARS2	WDR92
CDK2AP2	LAIR1	PAX8	WRB
CEACAM21	LBH	PCGF6	XCL1
CEBPE	LCE1E	PCYOX1	XLOC_000787
CENPBD1	LIMA1	PDE4DIP	XLOC_007052
CENPJ	LINC00324	PDE7A	XLOC_011219
CENPK	lincRNA:chr1:113540052-113551227_F	PDGFB	XLOC_013282
CENPQ	lincRNA:chr1:147706526-147727651_R	PGBD2	XLOC_013283
CEP19	lincRNA:chr1:201482952-201545677_F	PHF13	XLOC_013837
CEP250	lincRNA:chr1:246948452-246962677_F	PHF23	XPO5
CEP76	lincRNA:chr10:29699919-29733919_F	PIGV	YIPF5
CETN3	lincRNA:chr10:88688970-88694220_F	PIN4	ZCCHC6
CGN	lincRNA:chr10:89908970-89919320_R	PIP5K1B	ZFP64
CHCHD7	lincRNA:chr11:6764642-6767690_R	PLEKHO2	ZGLP1
CIRBP	lincRNA:chr11:82783427-82803502_F	PNP	ZKSCAN3
CISD1	lincRNA:chr13:31179400-31191300_R	PNPLA7	ZMYM6NB
CISH	lincRNA:chr15:69373209-69387524_F	POLH	ZNF137P
CLCN3	lincRNA:chr17:44271698-44279998_F	PPM1D	ZNF148
CLN5	lincRNA:chr17:44271698-44280036_F	PPP1R10	ZNF167
CMKLR1	lincRNA:chr17:45149116-45150265_R	PRDX3	ZNF17
CNIH4	lincRNA:chr18:10552875-10561225_R	PREPL	ZNF175
COLEC11	lincRNA:chr19:28266143-28266711_R	PRIM1	ZNF189
COPS5	lincRNA:chr19:28266664-28268543_R	PRPSAP1	ZNF2
CPNE1	lincRNA:chr19:28269395-28283939_R	PRSS16	ZNF202
CRAT	lincRNA:chr19:28269547-28283939_R	PRSS36	ZNF223
CREBL2	lincRNA:chr19:28293532-28294380_F	PSMA8	ZNF226
CRK	lincRNA:chr19:28296773-28297062_F	PSTPIP2	ZNF232
CROCCP3	lincRNA:chr19:37493217-37493673_F	PURA	ZNF234
CRYZ	lincRNA:chr19:56807196-56809238_F	RAB10	ZNF235
CTH	lincRNA:chr2:121054680-121065605_F	RAB22A	ZNF248
CTU1	lincRNA:chr2:168402979-168414804_F	RAD50	ZNF257
CU677518	lincRNA:chr2:32237771-32245946_F	RAD51D	ZNF260
CXCR7	lincRNA:chr2:70275822-70282955_R	RAD54B	ZNF283

Gene symbol			
CYB5B	lincRNA:chr22:25408700-25464025_R	RALA	ZNF284
CYHR1	lincRNA:chr3:114043485-114052926_F	RALBP1	ZNF285
DARS2	lincRNA:chr4:103337452-103350952_R	RAP2B	ZNF30
DB462629	lincRNA:chr4:15657202-15695627_F	RAVER2	ZNF322
DCAF7	lincRNA:chr4:28383262-28404481_R	RBM14	ZNF347
DCLRE1A	lincRNA:chr5:141839041-141871766_R	RBM4B	ZNF397
DDX58	lincRNA:chr5:44826637-44828692_F	RFESD	ZNF404
DENND2D	lincRNA:chr5:44828220-44828690_F	RGL1	ZNF420
DET1	lincRNA:chr5:86109465-86540020_F	RGS14	ZNF470
DGCR14	lincRNA:chr6:30260107-30262700_F	RGS19	ZNF48
DGKQ	lincRNA:chr6:30271847-30274737_R	RIMBP3	ZNF485
DHFRL1	lincRNA:chr6:30271849-30272880_R	RINT1	ZNF530
DIRC1	lincRNA:chr6:32862495-32869036_F	RLN1	ZNF557
DNAJC19	lincRNA:chr6:32862560-32869332_F	RLN2	ZNF567
DNAJC24	lincRNA:chr6:43988722-43993947_F	RMI1	ZNF571
DNM1P46	lincRNA:chr6:5013951-5056876_R	RNASEL	ZNF594
DNMT3B	lincRNA:chr6:86386083-86388435_F	RNF2	ZNF613
DOK1	lincRNA:chr6:86386801-86387887_F	RNF5	ZNF641
DOK4	lincRNA:chr7:105564174-105564567_R	RNF8	ZNF667
DOLPP1	lincRNA:chr7:112756783-112757127_R	RPA2	ZNF681
DPF2	lincRNA:chr7:17719170-17720283_R	RSG1	ZNF691
DYNLL1	lincRNA:chr7:77313167-77314643_R	RTN3	ZNF70
ECE2	lincRNA:chr8:102185399-102185979_R	RTN4R	ZNF71
EFNA4	lincRNA:chr8:102193085-102195404_R	SAC3D1	ZNF773
EGLN2	lincRNA:chr8:38616218-38630243_R	SAP18	ZNF775
EGR1	lincRNA:chr8:90623517-90627537_R	SAR1B	ZNF785
EIF5B	lincRNA:chr8:92071874-92082224_R	SARDH	ZNF786
ENST00000390450	lincRNA:chr9:123606154-123611279_F	SCG3	ZNF793
ENST00000390463	lincRNA:chr9:5610300-5622250_R	SENP8	ZNF816
ENST00000399269	lincRNA:chrX:102024544-102185719_R	SETX	ZNF84
ENST00000411846	lincRNA:chrX:130864569-130871024_R	SGK223	ZNF862
ENST00000416673	lincRNA:chrX:15690954-15744304_R	SH3BP5L	ZNF879
ENST00000427872	lincRNA:chrX:64628047-64628361_R	SH3PXD2A	ZNF880
ENST00000434336	lincRNA:chrX:6745350-6897125_F	SIRT4	ZNRD1-AS1
ENST00000444043	lincRNA:chrX:68406650-68432200_F	SKIV2L2	ZSCAN20
ENST00000450721	lincRNA:chrX:73164675-73232924_F	SKP2	ZSWIM3
ENST00000485364	LMAN2L	SLAIN1	ZW10
EPG5	LOC100128737	SLC16A10	
EPHA1	LOC100129138	SLC22A5	

Table 7.13: Genes upregulated in SLE total Treg

Gene symbol			
A_24_P49597	EPSTI1	IFI44L	OAS3
AHNAK	ERMN	IFI6	ODF3B
APOBEC3A	FAM183B	IL1RN	OR52N4

Gene symbol			
APOBEC3B	FLJ22447	LAMP3	PER1
BCL3	FLJ45684	LINC00487	PLBD1
CCDC157	FOSL2	lincRNA:chr2:54284416-54307601_R	PLSCR1
CCDC19	G0S2	LOC100509175	RPS26
CCR1	GRASP	LOC254057	RSAD2
CMPK2	GRM2	MAFB	SETD7
CR605633	HERC5	MAP3K8	SLC3A1
EDN1	IFI27	METRNL	USP18
EIF2AK2	IFI44	NR4A3	VAV3

Table 7.14: Genes downregulated in SLE total Treg

Gene symbol			
A_33_P3398196	FAM43A	lincRNA:chr10:88688970-88694220_F	SLC35F3
ARHGAP42	GOLPH3L	lincRNA:chr18:10552875-10561225_R	YIPF5
ATG4C	HKDC1	lincRNA:chr19:28293532-28294380_F	ZFP14
CCDC126	HOMEZ	lincRNA:chr2:121054680-121065605_F	ZMYM6NB
CD200R1	IDI2-AS1	lincRNA:chr2:32237771-32245946_F	ZNF175
CENPQ	ITGA10	lincRNA:chr6:108346507-108355057_R	ZNF232
CXCR7	LCE1E	LOC100131138	ZNF322
ENST00000485364	LIMA1	RIMBP3	ZNF404

Table 7.15: Genes upregulated in SLE total Tresp

Gene symbol			
A_24_P49597	ERMN	LOC100509175	RAMP1
ANK3	EXT1	LOC254057	RPS26
APOBEC3A	GADD45A	LOC643770	SLC16A6
C10orf68	GHRL	MAX	SLC3A1
CCDC19	HERC5	MED23	TMEM2
CCT6B	IFI27	NFKBIE	TOX
CD83	IFI44L	ODF3B	UPB1
CFL2	IL20RB	PCGF5	USP41
CPLX3	KLF11	PDE4DIP	VAV3
DST	KLHL15	PELI1	XM_003118559
EIF2AK2	LAMP3	PER1	ZC3H12A
EPSTI1	lincRNA:chr2:54284416-54307601_R	PIM3	ZRSR2

Table 7.16: Genes downregulated in SLE total Tresp

Gene symbol			
A_33_P3282213	FANCE	lincRNA:chr8:92071874-92082224_R	THAP3
A_33_P3398196	FBLN5	lincRNA:chrX:64628047-64628361_R	THAP8
AK124695	FCRLB	LOC100131131	THNSL1
ANKRD58	FIGNL1	LOC100131138	TMEM187
ARFIP1	FLJ39639	LOC100132057	TMEM60

Gene symbol			
ARHGAP42	FSD1	LOC100507199	TMEM8C
ATG4C	GOLPH3L	LOC728705	TRMT1L
BC039426	HKDC1	LRRC27	UBE2T
BPTF	HN1L	MAB21L2	UBTD2
C9orf69	ITGA10	METTL13	XLOC_013282
CENPBD1	JRKL	MRPS14	ZNF175
CENPQ	KLHL9	MYOM2	ZNF2
CEP19	LAIR1	NAPEPLD	ZNF226
CISH	LCE1E	NUP43	ZNF232
COLEC11	LIMA1	OSBPL11	ZNF260
CXCR7	lincRNA:chr1:113540052-113551227_F	PDGFB	ZNF322
CYHR1	lincRNA:chr10:89908970-89919320_R	RAD54B	ZNF404
DB462629	lincRNA:chr2:121054680-121065605_F	RIMBP3	ZNF470
DGCR14	lincRNA:chr2:168402979-168414804_F	RLN1	ZNF530
DHFRL1	lincRNA:chr22:25408700-25464025_R	RLN2	ZNF594
ENST00000390463	lincRNA:chr7:112756783-112757127_R	SLC35F3	ZNF613
ENST00000399269	lincRNA:chr8:38616218-38630243_R	SPATA21	ZNF793
ENST00000485364	lincRNA:chr8:90623517-90627537_R	STT3A	

Table 7.17: Genes upregulated in SLE Treg naive

Gene symbol
Zero transcripts identified

Table 7.18: Genes downregulated in SLE Treg naive

Gene symbol
CD200R1

Table 7.19: Genes upregulated in SLE Treg memory

Gene symbol			
APOBEC3A	CD163	LINC00487	MAFB
CCR1	IFI27	LOC100509175	

Table 7.20: Genes downregulated in SLE Treg memory

Gene symbol
Zero transcripts identified

Table 7.21: Genes upregulated in SLE Tresp naïve

Gene symbol		
EPSTI1	OAS3	CPLX3

Table 7.22: Genes downregulated in SLE Tresp naïve

Gene symbol
Zero transcripts identified

Table 7.23: Genes upregulated in SLE Tresp memory

Gene symbol
IGFL2

Table 7.24: Genes downregulated in SLE Tresp memory

Gene symbol			
DGCR14	EOMES	FASLG	ZNF175

Table 7.25: Non-coding transcripts upregulated in Treg

			Treg subset of upregulation		
Probe ID	***Accession***	***Gene name***	***Naïve Treg***	***Memory Treg***	***Total Treg***
A_32_P8813	LOC283663	uncharacterised LOC283663	✓		
A_19_P00316450	lincRNA:chr3:8615882-8616354_F	lincRNA:chr3:8615882-8616354_F		✓	
A_23_P170378	PMCHL1	pro-melanin-concentrating hormone-like 1, pseudogene		✓	
A_23_P166779	LINC00312	long intergenic non-protein coding RNA 312		✓	
A_32_P42725	ENST00000432978	ENST00000432978			✓
A_33_P3846114	LOC286442	uncharacterised LOC286442			✓
A_33_P3718269	ENST00000517927	ENST00000517927			✓
A_19_P00317691	lincRNA:chr13:50614546-50615364_R	lincRNA:chr13:50614546-50615364_R			✓
A_33_P3332616	A_33_P3332616	A_33_P3332616			✓
A_33_P3348519	ENST00000502284	ENST00000502284			✓
A_19_P00802201	lincRNA:chr7:7928975-7996750_R	lincRNA:chr7:7928975-7996750_R			✓
A_23_P60016	PTTG3P	pituitary tumor-transforming 3, pseudogene			✓
A_33_P3221613	ENST00000444919	ENST00000444919			✓
A_24_P145009	ZNRF2P1	zinc and ring finger 2 pseudogene 1			✓
A_33_P3357097	LOC100132741	uncharacterised LOC100132741			✓
A_33_P3254380	SLC9A7P1	solute carrier family 9 (sodium/hydrogen exchanger), member 7 pseudogene 1			✓
A_32_P104063	CRNDE	colorectal neoplasia differentially expressed (non-protein coding)			✓
A_19_P00331075	lincRNA:chr3:106619835-106630535_R	lincRNA:chr3:106619835-106630535_R			✓
A_19_P00316108	lincRNA:chr2:64455559-64478475_F	lincRNA:chr2:64455559-64478475_F			✓
A_33_P3416142	LOC100131234	familial acute myelogenous leukemia related factor			✓
A_33_P3337609	ENST00000411475	ENST00000411475			✓
A_19_P00321651	lincRNA:chr2:213437296-213768199_R	lincRNA:chr2:213437296-213768199_R			✓

7.2 Donor information

Table 7.26: Details of healthy donors

Healthy donor age range is from 23 to 54 years with a median of 30 and standard deviation of 12 years

Donor ID	*Gender*	*Age (years)*	*Ethnicity*
HC1	Female	31	Caucasian
HC2	Female	23	Caucasian
HC3	Female	30	Caucasian
HC4	Female	30	Caucasian
HC5	Female	54	Caucasian

Table 7.27: Details of SLE donors

				Clinical details at time of sampling				
Donor ID	*Gender*	*Age*	*Ethnicity*	*Disease severity*	*C3 (g/L)*	*Anti-dsDNA antibody titre*	*Features*	*Treatment*
SLE1	Female	36	Caucasian	Mild	Normal	Normal	Athritis, dermatological and neurological involvement	DMARDs (hydroxychloroquine), NSAIDs (aspirin), corticosteroids (prednisolone), gastroprotective agents (lansoprazole), calcium supplementation
SLE2	Female	21	Caucasian	Mild	Normal	Normal	Arthritis, neurological involvement	DMARDs (mycophenolate mofetil, hydroxychloroquine), NSAIDs (aspirin, piroxicam), calcium supplementation
SLE3	Female	37	Caucasian	Mild	Normal	Normal	General	No medication
SLE4	Female	32	Caucasian	Moderately active	0.82	97	Arthritis, fatigue	DMARDs (hydroxychloroquine), NSAIDs (aspirin, piroxicam), corticosteroids (prednisolone)
SLE5	Female	34	Caucasian	Moderately active	0.58	Normal	Renal disease	DMARDs (hydroxychloroquine, azathioprine), corticosteroids (prednisolone)

7.3 Th1 and Th2 specific genes identified by microarray

Table 7.28: Genes upregulated in Th1 cells

Gene symbol					
lincRNA:chr12:68421607-68422081_F	IL15	KCNA2	KLHDC7B	OAS3	FAM172BP
CCR1	DOK5	lincRNA:chr2:111992304-111994467_R	SH3RF1	PELI2	C15orf53
G0S2	ENPP6	EIF2AK2	A_33_P3278684	C10orf10	GBP3
IFNG	ENST00000400702	CPLX1	GLUL	NP113779	LOC79015
ANXA3	BCL2L14	SH3BP5	lincRNA:chr1:205404014-205407007_R	IL23R	IFI30
C15orf48	EPHA4	IFI6	MX2	PRDM8	PLAC8
ITGA6	TLR1	NPDC1	EBF4	DDX60L	BC029255
IFI44L	CLIC2	RSAD2	ADCY4	ANTXR2	lincRNA:chr2:112005726-112123291_R
TMEM213	MTMR11	NAPSA	TM6SF1	PARP9	GCNT1
EPAS1	TSHZ3	LAG3	DPY19L4	FES	SYNM
ETV7	STX3	ENST00000426099	BICD1	SGPP2	XM_003118976
S100A9	CYB561	KAT2B	RASA3	lincRNA:chr1:89874262-89947412_F	CRISPLD1
PRR5L	IL18RAP	TSPAN15	P2RX5	SPTBN5	IL31RA
ALPK1	TMEM30B	BTBD11	FAM3C	EPSTI1	IFIH1
HSD11B1	DNAJC6	LILRB3	FAAH2	RORC	LAMP3
STEAP1	C4orf26	EBI3	RAMP1	GIMAP8	NR3C2
LOC100507421	ENST00000378350	lincRNA:chr11:36494424-36500099_F	MANEAL	DDX60	CXCL10
ABCA1	CCNA1	PARP12	lincRNA:chr22:25588075-25594350_F	ANK3	MYH6
CD38	CPLX3	MAMLD1	CEACAM1	CCNI2	PLEKHG3
ENST00000360485	EGFL6	CD8B	RARRES3	GBP4	UNQ6494
MYH7	CHST15	DNAJC12	PREX1	DTX3L	ATP9A
IL2	IL12RB2	EIF4E3	ATP2B4	PDGFRB	AF363068

Gene symbol					
PLEK	NPTX1	PRF1	lincRNA:chr4:8359032-8359781_R	PRG4	TNFSF13B
IL21	MX1	TNFRSF18	SSBP3	JUN	LOC100289580
ENPP2	LOC100128019	TMEM200A	LPGAT1	GPR174	MAP6
TIMD4	PNMA2	LAMB3	SMAD3	IL26	PARP14
TXK	RASGRF2	PITPNC1	MGC70870	SNCA	NELL2
CCR5	IRAK2	CDC42EP1	MICA	C4BPB	GPR155
ATP8B4	ELOVL7	A_24_P143653	CPNE8	PLSCR1	IL18R1
TRPS1	IFIT5	TRIM22	TNFSF10	CD274	LOC541471
XAF1	IRAK3	ITGA3	CXCL16	HOXB6	lincRNA:chr2:111965652-112101793_R
RIPK2	USP18	SAMD9	IRF1	AIRE	SOD2
ATP6V0A4	MBL1P	HBEGF	lincRNA:chr4:8357036-8359103_R	TPST2	C12orf35
GPR34	TSHZ2	NUAK2	EDA	OSBP2	ASPHD2
MATN4	DGKG	GALNT10	CD59	RAPGEF2	BBS12
TBX21	GADD45B	PPARG	ARNTL	HOXB3	STAT2
MAP3K5	LOC100506342	MAP3K8	PPP2R2B	OR52N4	PRIC285
CDO1	ZG16B	HES1	IL13RA1	BC026731	LOC100131733
IFI44	PCSK5	KBTBD11	ISG15	F2RL1	ISG20
FAM26F	SAMD9L	lincRNA:chr4:8357101-8357609_R	CAST	TRANK1	ARRDC3
SLAMF7	lincRNA:chr1:89874262-89947412_R	CLIC5	TG	HERC6	ITGAX
ALOX5	HS6ST2	CDYL2	PALLD	TCF7	TNFAIP2
HPGD	UBD	TIFA	GAS8	GBP1	CA6
IFIT1	RCAN2	FMOD	LOC100505719	BCL2L11	TTC39C
LYN	USP41	KIRREL2	TGFBR1	HOXB4	lincRNA:chr14:101297417-101297818_F
SHISA2	lincRNA:chr21:29817053-30047175_R	ELOVL6	RNF157	PVRL3	HLA-C

Gene symbol					
lincRNA:chr21:29818746-30047176_R	INSC	lincRNA:chr13:44760050-44767325_R	CARD16	ELOVL4	lincRNA:chr8:107055124-107072278_F
DMD	MLLT4	D4S234E	KRT33A	PYGM	GCH1
ENST00000390252	GADD45G	lincRNA:chr2:111954496-112125102_R	CARD17	PAG1	SYNE2
GREM2	CFH	CCL25	A_33_P3314091	STOM	RNF213
DAPK2	PACSIN1	CYFIP1	ZBTB47	HERC5	VOPP1
IFIT3	lincRNA:chr2:111994580-111996624_R	HIC1	DPF3	OTUD1	MUC1
PDE4B	lincRNA:chr2:111960680-111962218_R	EMR1	XLOC_003385	lincRNA:chr2:37325696-37331035_R	FCHSD2
LMO4	DOCK5	GHRL	RIMS3	MDM1	OAS2
NLRP7	CFHR3	CD40	HOXB2	PDE1B	DAZAP1
GNLY	MAST4	lincRNA:chr2:111964707-111965184_R	KIAA1671	lincRNA:chr17:78306364-78308071_R	HLA-B
ENST00000492167	LACC1	SLC41A2	DGCR14	A_33_P3410019	DYRK2
lincRNA:chr10:5156300-5171275_R	STYK1	KIAA1217	RHOBTB2	LOC729603	FAS
lincRNA:chr12:68383224-68415107_F	RORA	FEZ1	ENST00000390398	AHNAK	
HAVCR1	RHOU	ZDHHC19	CLEC2B	KCNA3	
SIPA1L2	ENST00000390247	RASGEF1A	LY6E	STT3B	
XLOC_013923	IFI35	CASP1	TDRD7	lincRNA:chr16:80631802-80632239_R	

Table 7.29: Genes upregulated in Th2 cells

Gene symbol					
CRH	SEMA5A	CCR8	CACNB2	NEK6	MYH10
KRT1	KCNK1	IGFBP2	RAC3	GPR125	GALNT3
LOC728218	LOC729970	CLDND2	MFSD4	lincRNA:chr6:22191803-22194614_F	LOC730091
MAOA	PECAM1	ZNF239	RGS9	HECW2	AK096239
PTGIS	LOC643650	TCEA3	C18orf56	PI16	HIP1
CALD1	FAM174B	C9orf47	lincRNA:chr7:129401614-129418239_R	CD244	GNB4
FLJ45983	RBM47	ENST00000513415	ENST00000399967	AW002507	MRAP2
HOMER2	LOR	ACTN3	ABCG2	GPR183	PSMA8
IL17RB	GJB2	lincRNA:chr18:45332602-45354102_F	RNF215	DACT1	lincRNA:chr14:96507172-96661947_F
TRIM49L2	VSIG10L	UCHL1	NEK10	SIGLEC12	ENST00000338857
KIAA1462	FNBP1L	ZP1	lincRNA:chr8:32785498-32902078_F	TMEM98	C14orf132
HTR2B	PCDH8	ANKRD57	EEF2K	lincRNA:chr3:171506056-171528740_R	LOC100134285
TRIM43	lincRNA:chr3:171506370-171528740_R	ENST00000460965	lincRNA:chr8:104254399-104295074_R	OGDHL	NMUR1
CR604707	SIGLECP3	FRRS1	METTL9	LOC100506220	ACCN2
RNF43	C8orf46	CTDSPL	lincRNA:chr14:71276672-71284022_F	lincRNA:chr12:4958789-4970589_F	OSBPL1A
LRRC32	lincRNA:chr8:37258592-37350868_R	TPTE2P6	WFDC1	BCAR3	MBOAT1
lincRNA:chr1:214512902-214523621_R	lincRNA:chr6:138961057-139020207_R	PACSIN3	C1orf162	CA2	MTL5
C15orf27	SPINK2	CPNE5	CFD	lincRNA:chr7:105517604-105553086_F	DYNC2LI1
MAOB	LPCAT2	PDE9A	C6orf99	LOC286442	TLE2
PTPN20B	SCPEP1	FRG2	GPER	PLEKHH2	NSMCE1
LIMA1	LOC728084	AUH	GJC1	ITM2C	LOC400684
S1PR3	AF074986	ARHGEF26	MRPS26	SIGLEC10	ZNF488
PDE7B	XLOC_013837	GATM	ST7	IFNGR2	LOC730101

Gene symbol					
LOC100129034	LOC728819	TERT	SEC11C	HPGDS	COMP
ANXA8L2	SOX4	ENST00000378416	CD33	A_33_P3250273	lincRNA:chr2:232366756-232372381_R
ENST00000549427	ZBTB8A	A_33_P3333985	TMPRSS13	ZNF697	HILPDA
RXRA	PLD1	MLC1	IQCJ-SCHIP1	PIWIL2	SCHIP1
PLA2G4A	NRP1	lincRNA:chr6:22146709-22147336_R	LMCD1	RRS1	FAM171A1
CLIC6	PPP1R14A	lincRNA:chr8:58106178-58152853_R	lincRNA:chr2:217380530-217421230_F	lincRNA:chr7:105525814-105538089_F	AKAP12
THC2685600	ABHD6	NPAS2	TGFBR3	SPINT2	SLC40A1
PTPN20A	PNMAL1	ABCC4	CEP70	DUSP6	CCR4
THC2675356	FAM124B	PLAU	RUNX2	ADORA2B	SLC45A3
TPRG1	REEP2	B3GAT1	GALNTL4	C9orf135	OXER1
C8orf85	SLC39A14	SLC17A3	ARG2	SVOPL	C10orf55
LOC100505806	SGIP1	lincRNA:chr6:5013951-5056876_F	MARCKS	UTS2	TRAM2
AK075182	ERRFI1	PTPRN2	lincRNA:chr16:54944999-54959642_R	SMAD1	ACTN1
CAPSL	NM_144710	XDH	TTN	RYR1	LOC728671
FBLN5	SLC47A1	XLOC_006828	SCAI	ADHFE1	LINC00263
MYO3B	LINC00340	OXCT2	AGBL2	lincRNA:chr2:224501706-224587681_R	SGSH
IGF2BP3	CYP19A1	IL12A	HIST1H1A	SMAD2	lincRNA:chr8:32785489-32902343_F
CLDN1	ENST00000536986	RNF125	lincRNA:chr10:97639610-97667003_F	CDON	LOC727916
A_33_P3295623	lincRNA:chr7:105551484-105564551_F	PTPRVP	LOC100506105	SLMO1	CTTN
PTGDR2	ECSCR	LOC100505933	ENPP3	NFIL3	lincRNA:chr2:39745746-39826668_F
CHDH	ZYG11A	INPP4B	LOC442132	ZMAT1	GAMT
ARMCX1	lincRNA:chr1:101518537-101561912_F	BFSP2	TLR5	CYSLTR2	HSBP1L1
MRC2	SPATA3	GSTA4	ETS2	AGPAT9	DFNB59

Gene symbol					
AOAH	CTNNAL1	C21orf96	LINC00239	NUDT9P1	TRPC1
lincRNA:chr8:67304071-67319246_R	DNAJC5B	TRIM2	PTER	TRIO	TMEM45B
MEOX1	IL9	FLJ21408	HS3ST1	THC2680609	SPRYD7
ENPP1	FAM150B	PIP5K1B	CNR2	GLT1D1	PEX2
TNFSF11	THY1	XLOC_005143	SNTA1	XM_942053	A_33_P3390939
BHLHE22	A_24_P229726	SLC9A7P1	P2RX1	AHRR	MIR600HG
ENST00000442072	MGST3	SH3BGR	B9D1	DEPDC4	LIMS2
FZD3	AXIN2	CR591103	PNPLA3	LOC100132774	ALDH5A1
CABLES1	A_33_P3416196	NME4	ENST00000309775	EPN2	BC039426
DDC	THBS4	ACSM3	SCARB1	BC039440	TEX14
lincRNA:chr5:9547425-9554550_F	CERCAM	GATA3	ETFB	RGS12	PCSK4
GJB6	IL31	AGAP7	HSPA4L	ZNF532	SLC15A3
CDCP1	CACNA1D	LOC388780	GNRH2	AR	lincRNA:chrX:57003000-57010915_R
MYL9	BMS1P1	TMEM44	lincRNA:chr1:101548009-101552817_F	SORD	MAP2K6
PMCH	GDF10	RNASE4	NT5DC2	lincRNA:chr14:96555696-96556712_F	BARX1
GAD1	ENST00000423719	GML	ST7-AS1	ISYNA1	SH3RF2
IKZF2	LOC400768	SLC17A9	LRRC20	NKD2	PON2
ICAM4	STON1	BEX1	lincRNA:chr3:37864071-37876621_R	GAB2	S100P
PTPN14	CTHRC1	FV367791	SOX12	C10orf128	ANG
HRH4	RCBTB2	LOC642826	DIRC3	LOC100505683	SLIT3
GPR55	lincRNA:chr2:74193717-74210392_R	lincRNA:chr19:53702413-53709013_F	SYPL1	SLC4A7	FAM8A1
ENST00000431083	RUNX1	XAGE3	lincRNA:chr2:65103471-65188448_R	ACOXL	ZAR1
PMCHL1	C5	lincRNA:chr6:5047814-5048418_F	CLEC11A	lincRNA:chr4:128758375-128769050_F	C6orf81
ENST00000543306	lincRNA:chr14:96548697-96561747_F	lincRNA:chr2:208360030-208378555_R	AEBP1	ZFR2	FOXP1

Gene symbol					
NMT2	FLJ37798	PTPLA	CCDC151	GNAI1	PRKCA
NCS1	TBC1D4	lincRNA:chr10:97639610-97666960_R	ZNF827	KLHDC9	BEND5
TOX	RAB27B	C6orf52	NUDT11	ALDH7A1	LOC286063
MYBL1	LOC728558	PNOC	FLJ35946	NUDT7	
lincRNA:chr17:67547498-67549996_F	LOC348761	CAMK1	ENST00000445310	NDFIP2	
ATP12A	lincRNA:chr7:105551628-105553084_F	NUDT8	lincRNA:chr8:32853730-32902348_F	ACCS	
F12	ARMCX4	CEP55	GNAS-AS1	C1orf228	

7.4 T-bet super-enhancer eRNAs upregulated in Th1 cells

Table 7.30: T-bet super-enhancer eRNAs upregulated in Th1 cells

Comparison of super-enhancer eRNA FPKM values in Th1 and Th2 cells. RS and US conditions are also given. Super-enhancer IDs correspond to the nearest annotated genic TSS. Note that some super-enhancers correspond to more than one gene and some genes have more than one super-enhancer (named A, B). Transcripts within each super-enhancer are numbered individuallyAll are donor 2. FPKM: Fragments per kilbase per million mapped reads; RS: restimulated; US: unstimulated.

Super-enhancer		*eRNA*			*FPKM*								
Nearest gene	***Coordinates***	***ID***	***Coordinates***	***Strand***	***Th1 RS Total RNA***	***Naïve NA polyA***	***Th1 US polyA***	***Th1 RS polyA***	***Th2 US polyA***	***Th2 RS polyA***	***Th1 US Total RNA***	***Th2 US Total RNA***	***Th2 RS Total RNA***
ACSL6	chr5:131355565-131362793	ACSL6-eRNA	chr5:131356657-131357166	+	2.3055	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA1	chr5:157023417-157023721	-	2.7278	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA2	chr5:157023979-157024554	+	2.4523	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA3	chr5:157023706-157023927	+	3.7953	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA4	chr5:157026788-157027098	-	3.3749	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA5	chr5:157027429-157028493	+	4.7502	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA6	chr5:157029978-157030311	+	3.0170	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA7	chr5:157029992-157030455	-	1.8950	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA8	chr5:157028565-157029066	+	2.0414	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA9	chr5:157028654-157029833	-	2.2948	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA10	chr5:157030661-157031333	+	3.6196	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA11	chr5:157035113-157035593	+	1.4985	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ADAM19	chr5:157021683-157035243	ADAM19-eRNA12	chr5:157033477-157034749	-	5.6473	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ARL4C	chr2:235187010-235201525	ARL4C-eRNA1	chr2:235198404-235198555	-	5.5206	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ARL4C	chr2:235187010-235201525	ARL4C-eRNA2	chr2:235198636-235199094	+	4.6265	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ARL4C	chr2:235187010-235201525	ARL4C-eRNA3	chr2:235199199-235200084	+	1.9239	0.0000	0.0000	0.0000	0.0000	0.0000	1.6774	0.0000	0.5581

Super-enhancer		*eRNA*			*FPKM*								
Nearest gene	***Coordinates***	***ID***	***Coordinates***	***Strand***	***Th1 RS Total RNA***	***Naïve NA polyA***	***Th1 US polyA***	***Th1 RS polyA***	***Th2 US polyA***	***Th2 RS polyA***	***Th1 US Total RNA***	***Th2 US Total RNA***	***Th2 RS Total RNA***
BCL2	chr18:60867075-60888556	BCL2-eRNA2	chr18:60874451-60875137	+	1.8059	0.0000	2.5765	0.0000	4.9736	2.7211	1.1531	2.4339	1.7132
BHLHE40	chr3:5018681-5028927	BHLHE40-eRNA1	chr3:5018251-5018805	-	2.1900	0.0000	0.0000	0.0000	1.5204	1.6852	1.5601	0.0000	0.0000
BHLHE40	chr3:5018681-5028927	BHLHE40-eRNA2	chr3:5018911-5020038	-	3.7477	0.0000	0.0000	1.1805	1.3014	1.2163	1.5601	1.3390	0.9337
BZRAP1-BZRAP1-AS1-MIR142-MIR4736	chr17:56404868-56422171	BZRAP1-BZRAP1-AS1-MIR142-MIR4736-eRNA2	chr17:56421372-56421856	-	2.0149	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
C14orf182	chr14:50465940-50469622	C14orf182-eRNA1	chr14:50466125-50466481	+	3.0147	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
C14orf182	chr14:50465940-50469622	C14orf182-eRNA2	chr14:50468239-50468594	+	2.7222	0.0000	1.4524	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CCND2_A	chr12:4251899-4263091	CCND2_A-eRNA	chr12:4255795-4256015	+	2.9997	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CCND2_B	chr12:4378376-4382287	CCND2_B-eRNA1	chr12:4379265-4379590	+	2.5179	0.0000	0.0000	1.9808	2.2714	1.8225	0.0000	0.0000	0.0000
CCND2_B	chr12:4378376-4382287	CCND2_B-eRNA2	chr12:4381542-4381780	+	2.4636	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CDC7	chr1:92014439-92022120	CDC7-eRNA	chr1:92014156-92014687	-	3.7775	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9815
CSF2	chr5:131434767-131447219	CSF2-eRNA1	chr5:131435172-131435732	-	1.3526	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CSF2	chr5:131434767-131447219	CSF2-eRNA2	chr5:131434302-131434947	-	2.3171	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CSF2	chr5:131434767-131447219	CSF2-eRNA3	chr5:131439647-131440231	+	0.8610	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CSF2	chr5:131434767-131447219	CSF2-eRNA4	chr5:131439735-131440156	-	1.4900	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CSF2	chr5:131434767-131447219	CSF2-eRNA5	chr5:131441469-131442119	+	1.6084	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CSF2	chr5:131434767-131447219	CSF2-eRNA6	chr5:131437391-131439434	-	2.5725	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CSF2	chr5:131434767-131447219	CSF2-eRNA7	chr5:131440368-131441404	+	2.2799	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CSF2	chr5:131434767-131447219	CSF2-eRNA8	chr5:131438074-131438922	+	7.0811	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CSF2	chr5:131434767-131447219	CSF2-eRNA9	chr5:131445138-131446685	-	6.6748	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Super-enhancer		*eRNA*			*FPKM*								
Nearest gene	***Coordinates***	***ID***	***Coordinates***	***Strand***	***Th1 RS Total RNA***	***Naïve NA polyA***	***Th1 US polyA***	***Th1 RS polyA***	***Th2 US polyA***	***Th2 RS polyA***	***Th1 US Total RNA***	***Th2 US Total RNA***	***Th2 RS Total RNA***
DAD1	chr14:23018961-23031423	DAD1-eRNA1	chr14:23029112-23029481	-	1.5900	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DAD1	chr14:23018961-23031423	DAD1-eRNA2	chr14:23029566-23030690	-	2.5197	0.0000	0.0000	0.0000	0.0000	0.0000	1.4594	0.0000	1.6159
DAD1	chr14:23018961-23031423	DAD1-eRNA3	chr14:23030947-23031474	+	1.1573	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DDIT4	chr10:74056869-74057959	DDIT4-eRNA	chr10:74056798-74057020	-	5.1205	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0450
DNAJC5B	chr8:66843934-66868440	DNAJC5B-eRNA1	chr8:66860843-66861343	+	1.4316	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DNAJC5B	chr8:66843934-66868440	DNAJC5B-eRNA2	chr8:66864987-66865346	+	1.6419	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DNAJC5B	chr8:66843934-66868440	DNAJC5B-eRNA3	chr8:66868189-66868713	+	1.3587	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DPP4	chr2:162809195-162810979	DPP4-eRNA2	chr2:162807513-162809532	-	5.2821	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DPP4	chr2:162809195-162810979	DPP4-eRNA3	chr2:162810486-162810797	-	2.3002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ETS1_A	chr11:128188606-128195137	ETS1_A-eRNA1	chr11:128193767-128194277	-	1.5003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9050
ETS1_A	chr11:128188606-128195137	ETS1_A-eRNA2	chr11:128194908-128195243	+	2.9147	1.5624	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ETS1_B	chr11:128159970-128175956	ETS1_B-eRNA1	chr11:128160890-128161198	+	2.3272	1.7242	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0655
ETS1_B	chr11:128159970-128175956	ETS1_B-eRNA2	chr11:128161301-128161787	+	1.6889	4.0460	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6368
ETS1_B	chr11:128159970-128175956	ETS1_B-eRNA3	chr11:128172067-128172569	-	4.6829	1.1694	0.0000	0.0000	0.0000	0.9912	2.9895	2.5651	2.4060
FAM3C	chr7:121170265-121171217	FAM3C-eRNA	chr7:121171080-121172242	+	3.8357	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
FASLG_A	chr1:172668245-172677990	FASLG_A-eRNA1	chr1:172667787-172669169	-	1.9274	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
FASLG_A	chr1:172668245-172677990	FASLG_A-eRNA2	chr1:172667799-172669328	+	1.8440	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
FASLG_A	chr1:172668245-172677990	FASLG_A-eRNA3	chr1:172673366-172673879	+	1.4906	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
FASLG_A	chr1:172668245-172677990	FASLG_A-eRNA4	chr1:172672704-172673047	-	2.2046	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Super-enhancer		*eRNA*			*FPKM*								
Nearest gene	***Coordinates***	***ID***	***Coordinates***	***Strand***	***Th1 RS Total RNA***	***Naïve NA polyA***	***Th1 US polyA***	***Th1 RS polyA***	***Th2 US polyA***	***Th2 RS polyA***	***Th1 US Total RNA***	***Th2 US Total RNA***	***Th2 RS Total RNA***
FASLG_A	chr1:172668245-172677990	FASLG_A-eRNA5	chr1:172677906-172678241	+	1.7812	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
FASLG_B	chr1:172631475-172644841	FASLG_B-eRNA1	chr1:172641155-172642259	+	1.7263	0.0000	0.0000	2.8497	0.0000	0.9179	0.0000	0.0000	0.0000
FASLG_B	chr1:172631475-172644841	FASLG_B-eRNA2	chr1:172638368-172640852	+	3.4000	0.0000	0.0000	1.2570	0.0000	4.2255	0.0000	0.0000	1.3290
FURIN	chr15:91395196-91400410	FURIN-eRNA1	chr15:91395070-91395587	+	1.5762	0.0000	2.8526	2.1995	0.0000	0.0000	1.5772	0.0000	0.0000
FURIN	chr15:91395196-91400410	FURIN-eRNA2	chr15:91395777-91395913	+	5.9337	0.0000	2.9839	0.0000	0.0000	0.0000	1.5772	0.0000	0.0000
FURIN	chr15:91395196-91400410	FURIN-eRNA3	chr15:91395099-91395937	-	2.5633	0.0000	0.0000	0.0000	0.0000	0.0000	1.3259	0.0000	0.0000
FURIN	chr15:91395196-91400410	FURIN-eRNA4	chr15:91398235-91398700	+	2.4404	0.0000	1.7486	1.4920	0.0000	0.0000	1.2746	0.0000	0.0000
FURIN	chr15:91395196-91400410	FURIN-eRNA5	chr15:91399760-91399908	-	7.1107	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GIMAP2	chr7:150360853-150374281	GIMAP2-eRNA	chr7:150365291-150365691	-	1.9091	0.0000	0.0000	0.0000	0.0000	0.0000	1.9046	0.0000	0.0000
GZMB	chr14:25134509-25143870	GZMB-eRNA1	chr14:25134585-25134900	-	4.8789	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GZMB	chr14:25134509-25143870	GZMB-eRNA2	chr14:25135162-25135595	+	2.5253	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GZMB	chr14:25134509-25143870	GZMB-eRNA3	chr14:25139478-25139867	-	2.5833	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GZMB	chr14:25134509-25143870	GZMB-eRNA4	chr14:25140444-25141198	+	1.9570	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GZMB	chr14:25134509-25143870	GZMB-eRNA5	chr14:25141916-25142135	-	11.7962	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GZMB	chr14:25134509-25143870	GZMB-eRNA6	chr14:25142187-25142738	-	4.0391	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GZMB	chr14:25134509-25143870	GZMB-eRNA7	chr14:25143615-25145250	+	6.1301	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HAVCR1	chr5:156452061-156458115	HAVCR1-eRNA1	chr5:156452040-156452349	-	2.3181	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HAVCR1	chr5:156452061-156458115	HAVCR1-eRNA2	chr5:156455530-156455718	-	4.8793	2.0498	1.7704	1.6278	0.0000	0.0000	0.0000	0.0000	0.0000
HAVCR1	chr5:156452061-156458115	HAVCR1-eRNA3	chr5:156453808-156454390	-	1.1235	3.1777	3.2988	2.3066	0.0000	0.0000	0.9958	0.0000	0.0000
HAVCR1	chr5:156452061-156458115	HAVCR1-eRNA4	chr5:156457466-156457504	+	107.4261	0.0000	0.0000	0.0000	0.0000	0.0000	77.7862	97.7004	81.8291
IFNG	chr12:68557178-68584152	IFNG-eRNA1	chr12:68555552-68557747	-	6.5995	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Super-enhancer		eRNA			FPKM								
Nearest gene	Coordinates	ID	Coordinates	Strand	Th1 RS Total RNA	Naïve NA polyA	Th1 US polyA	Th1 RS polyA	Th2 US polyA	Th2 RS polyA	Th1 US Total RNA	Th2 US Total RNA	Th2 RS Total RNA
IFNG	chr12:68557178-68584152	IFNG-eRNA2	chr12:68557941-68558204	+	2.6064	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA3	chr12:68558346-68559126	+	3.0193	0.0000	0.0000	0.0000	0.0000	0.0000	1.6371	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA4	chr12:68560053-68560736	-	2.7581	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA5	chr12:68560996-68561819	+	1.0097	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA6	chr12:68563166-68564061	+	2.0912	0.0000	0.9318	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA7	chr12:68564507-68566330	-	1.0585	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA8	chr12:68565959-68566925	+	4.1601	0.0000	0.0000	0.0000	0.0000	0.0000	1.0490	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA9	chr12:68567631-68567788	-	4.7777	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA10	chr12:68567943-68570462	-	2.4613	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA11	chr12:68570568-68572497	-	2.0250	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA12	chr12:68573107-68574069	-	7.3884	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA13	chr12:68569397-68573651	+	4.7151	0.0000	1.0905	0.0000	0.0000	0.0000	1.3463	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA14	chr12:68574321-68575293	-	6.7932	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA15	chr12:68574930-68575825	+	3.5305	0.0000	0.0000	0.0000	0.0000	0.0000	0.8119	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA16	chr12:68575974-68576496	+	2.1930	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA17	chr12:68579518-68579839	+	4.2589	0.0000	0.0000	0.0000	0.0000	0.0000	1.1004	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA18	chr12:68582527-68582731	+	7.3817	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA19	chr12:68582792-68582974	+	4.2228	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA20	chr12:68583093-68583432	+	2.5547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IFNG	chr12:68557178-68584152	IFNG-eRNA21	chr12:68583731-68584878	+	6.4636	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IKZF1	chr7:50415329-50428541	IKZF1-eRNA3	chr7:50427621-50428323	-	1.1978	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Super-enhancer		*eRNA*			*FPKM*								
Nearest gene	***Coordinates***	***ID***	***Coordinates***	***Strand***	***Th1 RS Total RNA***	***Naïve NA polyA***	***Th1 US polyA***	***Th1 RS polyA***	***Th2 US polyA***	***Th2 RS polyA***	***Th1 US Total RNA***	***Th2 US Total RNA***	***Th2 RS Total RNA***
IL1RL1	chr2:102869642-102879068	IL1RL1-eRNA1	chr2:102876118-102876419	+	2.2086	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IL1RL1	chr2:102869642-102879068	IL1RL1-eRNA2	chr2:102874335-102875233	-	1.1366	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IL1RL1	chr2:102869642-102879068	IL1RL1-eRNA3	chr2:102878148-102878499	+	1.6858	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
KLF6	chr10:3847408-3852963	KLF6-eRNA1	chr10:3847347-3848142	+	1.3559	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
KLF6	chr10:3847408-3852963	KLF6-eRNA2	chr10:3851566-3851830	-	2.3779	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
KLF6	chr10:3847408-3852963	KLF6-eRNA3	chr10:3852219-3852637	-	1.8777	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC00501	chr3:177076401-177078107	LINC00501-eRNA	chr3:177077927-177078367	+	1.2989	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC00861	chr8:126942586-126963824	LINC00861-eRNA2	chr8:126960301-126960620	+	2.4919	0.0000	0.0000	0.0000	0.0000	0.0000	2.5605	0.0000	0.0000
LINC00892	chrX:135709455-135717895	LINC00892-eRNA1	chrX:135710066-135710736	+	1.2599	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC00892	chrX:135709455-135717895	LINC00892-eRNA2	chrX:135710197-135710695	-	1.0272	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC00892	chrX:135709455-135717895	LINC00892-eRNA3	chrX:135709744-135709971	+	2.8800	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC00892	chrX:135709455-135717895	LINC00892-eRNA4	chrX:135711408-135712084	+	2.2753	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC00892	chrX:135709455-135717895	LINC00892-eRNA5	chrX:135707035-135709677	-	3.7462	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC00892	chrX:135709455-135717895	LINC00892-eRNA6	chrX:135708509-135709692	+	4.0882	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC00892	chrX:135709455-135717895	LINC00892-eRNA7	chrX:135716883-135717234	-	3.0651	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC00892	chrX:135709455-135717895	LINC00892-eRNA8	chrX:135717454-135717772	+	2.4119	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC01136	chr1:203256122-203259100	LINC01136-eRNA1	chr1:203258872-203259654	+	2.6347	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC01136	chr1:203256122-203259100	LINC01136-eRNA2	chr1:203258938-203259627	-	3.8105	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LINC01136	chr1:203256122-203259100	LINC01136-eRNA3	chr1:203258166-203258725	-	3.7824	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MIR1206	chr8:128989835-129005802	MIR1206-eRNA2	chr8:128992174-128992591	-	2.0335	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MIR1206	chr8:128989835-129005802	MIR1206-eRNA5	chr8:128994061-128994636	-	1.7931	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Super-enhancer		*eRNA*			*FPKM*								
Nearest gene	***Coordinates***	***ID***	***Coordinates***	***Strand***	***Th1 RS Total RNA***	***Naïve NA polyA***	***Th1 US polyA***	***Th1 RS polyA***	***Th2 US polyA***	***Th2 RS polyA***	***Th1 US Total RNA***	***Th2 US Total RNA***	***Th2 RS Total RNA***
MIR1206	chr8:128989835-129005802	MIR1206-eRNA6	chr8:129002678-129003394	-	1.4484	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MIR1206	chr8:128989835-129005802	MIR1206-eRNA9	chr8:128999402-129000026	-	1.2775	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MIR1206	chr8:128989835-129005802	MIR1206-eRNA10	chr8:129003534-129005286	-	15.2530	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MIR6855	chr9:132647939-132651721	MIR6855-eRNA	chr9:132647694-132648130	-	3.6987	2.0606	0.0000	2.2898	1.3187	2.6857	0.0000	1.8109	1.8140
P2RX5	chr17:3597137-3616557	P2RX5-eRNA	chr17:3608728-3608935	-	3.5459	0.0000	0.0000	2.0603	0.0000	0.8903	0.0000	0.0000	0.0000
PDE4B	chr1:66740707-66754118	PDE4B-eRNA2	chr1:66748774-66749233	-	2.0262	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PMF1	chr1:156186128-156188127	PMF1-eRNA3	chr1:156187492-156187857	-	3.2207	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PRDM1	chr6:106256956-106263358	PRDM1-eRNA1	chr6:106259426-106260049	+	0.8824	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PRDM1	chr6:106256956-106263358	PRDM1-eRNA2	chr6:106262299-106262651	-	1.6802	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PRSS2	chr7:142505541-142507253	PRSS2-eRNA1	chr7:142506816-142507034	-	2.7598	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PRSS2	chr7:142505541-142507253	PRSS2-eRNA2	chr7:142504344-142506747	-	4.8010	0.0000	0.0000	0.0000	0.0000	0.0000	1.0954	0.0000	0.0000
PRSS2	chr7:142505541-142507253	PRSS2-eRNA3	chr7:142507150-142508606	+	7.5967	0.0000	1.9048	0.0000	0.0000	0.0000	6.6907	0.0000	1.4580
PRSS2	chr7:142505541-142507253	PRSS2-eRNA4	chr7:142505671-142506729	+	2.1565	0.0000	1.9757	0.0000	0.0000	0.6960	1.4467	0.9551	0.0000
PTGER4	chr5:40428252-40442140	PTGER4-eRNA1	chr5:40436922-40437396	-	1.8455	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PTGER4	chr5:40428252-40442140	PTGER4-eRNA2	chr5:40439610-40440051	+	2.0022	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
RASA3	chr13:114907886-114914450	RASA3-eRNA	chr13:114913463-114914439	-	1.6847	1.2028	4.4353	0.0000	0.0000	0.0000	2.2126	0.0000	0.0000
RASGRP1	chr15:38903099-38903891	RASGRP1-eRNA	chr15:38903557-38903845	+	2.5245	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
RP11-162J8.2	chr6:149449610-149469908	RP11-162J8.2-eRNA	chr6:149460824-149461379	-	1.6392	0.0000	0.0000	1.2020	0.0000	0.0000	2.8108	1.2477	0.0000
RP11-402G3.3	chr9:117453186-117459756	RP11-402G3.3-eRNA1	chr9:117454418-117454876	+	2.1440	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
RP11-	chr9:1174531	RP11-	chr9:117455055	+	2.4969	0.0000	0.0000	1.1046	0.0000	0.0000	0.0000	0.0000	0.0000

Super-enhancer		*eRNA*			*FPKM*								
Nearest gene	***Coordinates***	***ID***	***Coordinates***	***Strand***	***Th1 RS Total RNA***	***Naïve NA polyA***	***Th1 US polyA***	***Th1 RS polyA***	***Th2 US polyA***	***Th2 RS polyA***	***Th1 US Total RNA***	***Th2 US Total RNA***	***Th2 RS Total RNA***
402G3.3	86-117459756	402G3.3-eRNA2	-117455529										
RP11-402G3.3	chr9:117453186-117459756	RP11-402G3.3-eRNA3	chr9:117453593-117454351	+	2.7243	0.0000	0.0000	0.0000	0.0000	0.0000	1.4632	0.0000	0.0000
RP11-402G3.3	chr9:117453186-117459756	RP11-402G3.3-eRNA4	chr9:117457340-117457613	+	3.7327	0.0000	0.0000	1.0779	0.0000	0.0000	0.0000	0.0000	0.0000
RP11-402G3.3	chr9:117453186-117459756	RP11-402G3.3-eRNA5	chr9:117459195-117459402	+	2.9549	0.0000	0.0000	1.6345	0.0000	0.0000	0.0000	0.0000	0.0000
RP11-402G3.3	chr9:117453186-117459756	RP11-402G3.3-eRNA6	chr9:117459048-117459360	-	1.7626	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
RUNX3	chr1:25278807-25293528	RUNX3-eRNA2	chr1:25292287-25293209	-	1.3745	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
RUNX3	chr1:25278807-25293528	RUNX3-eRNA3	chr1:25292400-25292992	+	0.9231	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
SATB1	chr3:18754261-18767707	SATB1-eRNA1	chr3:18754020-18754599	-	0.9561	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
SATB1	chr3:18754261-18767707	SATB1-eRNA6	chr3:18766081-18767084	-	1.9257	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
SLAMF7	chr1:160725755-160727039	SLAMF7-eRNA1	chr1:160726697-160727291	+	1.6480	0.0000	0.0000	1.2674	0.0000	0.0000	0.0000	0.0000	0.0000
SLAMF7	chr1:160725755-160727039	SLAMF7-eRNA2	chr1:160726014-160726627	+	2.4494	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
SLC39A10	chr2:196397023-196408137	SLC39A10-eRNA	chr2:196398283-196398880	-	1.5969	0.0000	0.0000	0.0000	0.0000	0.0000	1.3776	0.9374	0.8830
SNORA14B	chr1:235147261-235148139	SNORA14B-eRNA1	chr1:235147882-235148623	+	5.4502	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
STAT1	chr2:191884859-191915035	STAT1-eRNA1	chr2:191887037-191887457	-	1.3695	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
STAT1	chr2:191884859-191915035	STAT1-eRNA2	chr2:191893432-191893944	-	1.2448	1.3357	0.0000	0.0000	0.0000	0.0000	0.0000	1.5157	0.6829
TBKBP1	chr17:45780392-45798949	TBKBP1-eRNA1	chr17:45797233-45797694	+	1.3442	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TBKBP1	chr17:45780392-45798949	TBKBP1-eRNA2	chr17:45795790-45796638	+	4.0843	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TBKBP1	chr17:45780392-45798949	TBKBP1-eRNA3	chr17:45798812-45800365	+	5.8705	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TNFAIP8_B	chr5:118664508-118679814	TNFAIP8_B-eRNA3	chr5:118675331-118676206	-	1.1445	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Super-enhancer		*eRNA*			*FPKM*								
Nearest gene	***Coordinates***	***ID***	***Coordinates***	***Strand***	***Th1 RS Total RNA***	***Naïve NA polyA***	***Th1 US polyA***	***Th1 RS polyA***	***Th2 US polyA***	***Th2 RS polyA***	***Th1 US Total RNA***	***Th2 US Total RNA***	***Th2 RS Total RNA***
TNFAIP8_B	chr5:118664508-118679814	TNFAIP8_B-eRNA6	chr5:118678516-118679446	-	0.9485	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

7.5 Genes with similar expression patterns to Treg ncRNAs

Table 7.31: Treg relevant genes associated with *CRNDE* expression

Gene name					
A_33_P3241043	CDCA2	FAM164A	KIAA1841	NPC1	SMC4
A_33_P3333985	CDCA5	FAM58BP	KIF11	NUSAP1	SMC6
ABHD5	CDCA7	FAM72A	KIF14	OAS1	SNED1
ACTA2	CDCA8	FAM72D	KIF15	OAS2	SNORD3B-1
ADAM19	CDHR3	FAM83D	KIF18A	OBFC2A	SNX10
ALCAM	CDK1	FANCA	KIF23	OGN	SPATS2L
ANLN	CDKN3	FANCI	KLF11	OOEP	SPC25
ANXA2	CEBPA	FANCL	LAYN	OR5T3	SPSB1
ARHGAP11A	CENPA	FANK1	LINC00467	PARP12	SRGAP2
ASPM	CENPE	FHL2	lincRNA:chr1:211556096-211605878_F	PARVB	ST7-AS1
ATP1B1	CENPF	FLJ45482	lincRNA:chr10:97639610-97667003_F	PBK	ST8SIA4
ATP2B4	CENPN	FLJ45983	lincRNA:chr11:74358152-74366327_R	PCK2	STAM
ATXN7L1	CENPW	FLNA	lincRNA:chr12:4958789-4970589_F	PCNA	STIL
B3GNT5	CEP128	GAB2	lincRNA:chr16:80631802-80632239_R	PCTP	TBC1D2
BARD1	CEP55	GADD45A	lincRNA:chr2:111954496-112125102_R	PDE1B	TFRC
BATF	CHDH	GCC2	lincRNA:chr2:111960179-112010479_R	PHTF1	THC2477815
BATF3	CHEK1	GINS1	lincRNA:chr2:111965652-112101793_R	PLEKHH3	TJP2
BCAS1	CHRNA6	GINS2	lincRNA:chr2:111992304-111994467_R	PMAIP1	TK1
BFSP2	CHST11	GMNN	lincRNA:chr2:111997376-112123246_R	PMCH	TMEM38A
BIRC5	CIT	GNA15	lincRNA:chr2:112005726-112123291_R	PMCHL1	TNFRSF8
BLM	CKS2	GPR19	lincRNA:chr2:213703255-213723530_F	POLE2	TNFRSF9
BRCA1	CLSPN	GRAMD4	lincRNA:chr3:130569543-130575125_F	POLQ	TOP2A
BRCA2	CNTNAP1	GTF2IRD1	lincRNA:chr3:171506056-171528740_R	PPP1R26	TOX
BRIP1	CORO2A	GXYLT2	lincRNA:chr7:7928975-7996750_R	PPP1R2P9	TPM4
BTG3	CPNE2	HECW2	lincRNA:chrX:45897556-45923781_R	PTTG1	TPX2
BUB1	CREG2	HERPUD1	LMCD1	PTTG2	TRIB1
BUB1B	CRNDE	HIST1H1B	LMNA	PTTG3P	TRIM16L
BZRAP1	CTNNA1	HIST1H2AI	LMNB1	RAB11FIP1	TRIM32
C11orf80	CTNNAL1	HIST1H2AL	LOC115110	RAB11FIP5	TRIP13
C12orf48	CTSA	HIST2H3A	LOC541471	RACGAP1	TTC7B
C12orf75	CTSC	HJURP	LOXL1	RAD51	TTK
C15orf53	DAB2IP	HMGB3	LRP8	RAD54L	TTN
C18orf1	DCP1B	HMGB3P1	LRRC32	RAP1GDS1	TUFT1
C8orf85	DIAPH3	HMMR	MCM4	RBBP8	TXN
C9orf167	DLGAP5	HN1	MCM5	RGPD1	TXNRD2
CARD6	DPYSL2	HPGD	MELK	RGS1	TYMS
CASC5	DSCC1	HS3ST3B1	MEOX1	RHOXF1	UBE2C
CASP8	DTL	HSPA1A	METRNL	RPL39L	UHRF1
CCDC50	DUSP10	HTATIP2	MGST2	RTKN2	UTS2

Gene name					
CCNA2	DUSP4	HTR2B	MKI67	SAT1	VAV3
CCNB2	E2F2	ICA1	MLF1IP	SCG5	WBSCR27
CCNE2	E2F7	IL12RB2	MND1	SEMA3G	WDHD1
CCNG2	ELK3	IL1R1	MSC	SGMS1	WEE1
CCR10	ENTPD1	IL7	MSMO1	SKA3	WNT10A
CCR3	EXO1	INPP1	MTHFD2	SKAP2	XLOC_011950
CCR4	EZH2	IRF5	MYO5C	SLC16A1	XM_001130734
CCR8	EZR	JAZF1	NBEAL2	SLC1A1	XM_003119674
CCRL2	F5	JSRP1	NDC80	SLC2A8	ZNF193
CD59	FAM110A	KCNN4	NEK10	SLC35F2	ZWINT
CD70	FAM110C	KEAP1	NFE2L3	SLC7A5	
CDC45	FAM124B	KIAA0101	NINJ2	SMAD1	

Table 7.32: Treg relevant genes associated with *LOC286442* expression

Gene symbol					
A_33_P3324409	CR600327	GSTA4	lincRNA:chr13:40669725-40795050_F	PDE4D	TDRD9
A_33_P3339701	CTLA4	GTSE1	lincRNA:chr13:50613143-50613550_R	PER2	THC2495962
A_33_P3414022	CTNNAL1	HELLS	lincRNA:chr2:179914333-179915343_R	PML	TIRAP
ABCC4	CXorf21	HLA-DMA	lincRNA:chr2:200834655-200839730_F	POU2AF1	TMEM236
ACACB	CXorf64	HLA-DMB	lincRNA:chr2:27788931-27790011_R	PRLH	TMEM38A
ACSBG1	CYP3A4	HLA-DOA	lincRNA:chr3:114031960-114042235_F	PSMD9	TMEM44
AMZ1	DCHS2	HLA-DPA1	lincRNA:chr5:3612325-3623500_F	PTPLA	TMPPE
ANKRD54	DGCR11	HLA-DPB1	lincRNA:chr7:45019225-45026725_F	RAPH1	TNFRSF1B
APAF1	DISP2	HLA-DQB1	lincRNA:chr8:101925424-101930749_F	RBMS3	TP53RK
ARFIP1	DNAH8	HLA-DRA	lincRNA:chr8:103953859-103990104_F	REL	TPR
ARHGAP21	DPY19L3	HLA-DRB1	lincRNA:chr8:16113839-16424929_R	RTKN2	TRIO
BASP1	E2F1	HLA-DRB3	lincRNA:chr8:95917724-95933974_F	S100A7	TTN
BMF	ELK1	HLA-DRB4	lincRNA:chr8:96743021-96861105_R	SASS6	TTYH2
BMP1	ENST00000390247	HLA-DRB5	lincRNA:chrX:55936225-56241400_F	SEBOX	TWIST1
BMPR2	ENST00000432978	HUS1	LMBRD2	SELP	UAP1L1
BRCA1	ENST00000433110	HVCN1	LOC100130894	SEMA7A	UNC5C
BRCA2	ENST00000517927	ICA1	LOC100131234	SH3PXD2A	VDR
BRIP1	ENTPD1	IDI2-AS1	LOC100132741	SHMT2	VSIG10L
BTC	FAIM2	IKZF2	LOC143188	SLA2	XLOC_002613
C19orf71	FAM124B	IKZF4	LOC286442	SLC14A1	XLOC_007206
C22orf34	FAM19A2	IL1RAP	LOC392364	SLC15A1	XLOC_010376
C5orf30	FANCA	IL21	LOC651337	SLC15A4	XM_003118559
C6orf211	FANCB	INPP5F	MCU	SLC16A1	XM_003119219
CD200	FCRL1	IRF4	METTL7A	SLC43A2	YWHAE

Gene symbol					
CD302	FCRL2	KBTBD8	MLC1	SLC9A7	ZBTB32
CD40	FCRL3	KDM1B	MS4A6A	SLC9A7P1	ZNF100
CD74	FHL2	KLHL12	MXD3	SMPD3	ZNF185
CDK14	FOXP3	KLK1	MYO1E	SNAP29	ZNF684
CENPL	GEN1	KRTAP11-1	NEDD4L	SSC5D	ZNF80
CGB2	GFRA2	LAT2	OR5H6	STAM	ZWINT
CHRM5	GK	LINC00167	OR7E91P	STARD4	
CIITA	GLCCI1	lincRNA:chr10:72689994-72706119_R	OTUD4	STRADA	
CORO1C	GLCE	lincRNA:chr12:73263758-73528783_F	P2RX5	SWAP70	
CPPED1	GPM6A	lincRNA:chr12:76015433-76261508_R	P2RY1	TBX2	

8 References

1. Tada, T., T. Takemori, K. Okumura, et al., Two distinct types of helper T cells involved in the secondary antibody response: independent and synergistic effects of Ia- and Ia+ helper T cells. J Exp Med, 1978. 147(2): p. 446-58.
2. Mosmann, T.R., H. Cherwinski, M.W. Bond, et al., Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. J Immunol, 1986. 136(7): p. 2348-57.
3. Cherwinski, H.M., J.H. Schumacher, K.D. Brown, et al., Two types of mouse helper T cell clone. III. Further differences in lymphokine synthesis between Th1 and Th2 clones revealed by RNA hybridization, functionally monospecific bioassays, and monoclonal antibodies. J Exp Med, 1987. 166(5): p. 1229-44.
4. McKenzie, A.N., J.A. Culpepper, R. de Waal Malefyt, et al., Interleukin 13, a T-cell-derived cytokine that regulates human monocyte and B-cell function. Proc Natl Acad Sci U S A, 1993. 90(8): p. 3735-9.
5. de Waal Malefyt, R., C.G. Figdor, R. Huijbens, et al., Effects of IL-13 on phenotype, cytokine production, and cytotoxic function of human monocytes. Comparison with IL-4 and modulation by IFN-gamma or IL-10. J Immunol, 1993. 151(11): p. 6370-81.
6. Murphy, K., Travers, P., Walport, M., Ehrenstein, M., Mauri, C., Mowat, A., Shaw, A., Macrophage activation by Th1 cells, in Janeway's Immunobiology2008, Garland Science. p. 368-372.
7. Murphy, K., Travers, P., Walport, M., Ehrenstein, M., Mauri, C., Mowat, A., Shaw, A., B cell activation and antibody production, in Janeway's Immunobiology2008, Garland Science. p. 381-400.
8. Zhu, J. and W.E. Paul, Peripheral CD4+ T-cell differentiation regulated by networks of cytokines and transcription factors. Immunol Rev, 2010. 238(1): p. 247-62.
9. Murphy, K., Travers, P., Walport, M., Ehrenstein, M., Mauri, C., Mowat, A., Shaw, A., The generation of T-cell receptor ligands, in Janeway's Immunobiology2008, Garland Science. p. 182-195.
10. Heufler, C., F. Koch, U. Stanzl, et al., Interleukin-12 is produced by dendritic cells and mediates T helper 1 development as well as interferon-gamma production by T helper 1 cells. Eur J Immunol, 1996. 26(3): p. 659-68.
11. Hsieh, C.S., S.E. Macatonia, C.S. Tripp, et al., Development of TH1 CD4+ T cells through IL-12 produced by Listeria-induced macrophages. Science, 1993. 260(5107): p. 547-9.
12. Girdlestone, J. and M. Wing, Autocrine activation by interferon-gamma of STAT factors following T cell activation. Eur J Immunol, 1996. 26(3): p. 704-9.
13. Sokol, C.L., G.M. Barton, A.G. Farr, et al., A mechanism for the initiation of allergen-induced T helper type 2 responses. Nat Immunol, 2008. 9(3): p. 310-8.
14. Robb, R.J., A. Munck, and K.A. Smith, T cell growth factor receptors. Quantitation, specificity, and biological relevance. J Exp Med, 1981. 154(5): p. 1455-74.
15. Park, H., Z. Li, X.O. Yang, et al., A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. Nat Immunol, 2005. 6(11): p. 1133-41.
16. Bettelli, E., Y. Carrier, W. Gao, et al., Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. Nature, 2006. 441(7090): p. 235-8.
17. Chen, W., W. Jin, N. Hardegen, et al., Conversion of peripheral CD4+CD25- naive T cells to CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3. J Exp Med, 2003. 198(12): p. 1875-86.
18. Toda, A. and C.A. Piccirillo, Development and function of naturally occurring CD4+CD25+ regulatory T cells. J Leukoc Biol, 2006. 80(3): p. 458-70.
19. Pawankar, R., M. Okuda, H. Yssel, et al., Nasal mast cells in perennial allergic rhinitics exhibit increased expression of the Fc epsilonRI, CD40L, IL-4, and IL-13, and can induce IgE synthesis in B cells. J Clin Invest, 1997. 99(7): p. 1492-9.
20. Debray-Sachs, M., C. Carnaud, C. Boitard, et al., Prevention of diabetes in NOD mice treated with antibody to murine IFN gamma. J Autoimmun, 1991. 4(2): p. 237-48.
21. Kotake, S., N. Udagawa, N. Takahashi, et al., IL-17 in synovial fluids from

patients with rheumatoid arthritis is a potent stimulator of osteoclastogenesis. J Clin Invest, 1999. 103(9): p. 1345-52.
22. Cua, D.J., J. Sherlock, Y. Chen, et al., Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. Nature, 2003. 421(6924): p. 744-8.
23. Brunkow, M.E., E.W. Jeffery, K.A. Hjerrild, et al., Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. Nat Genet, 2001. 27(1): p. 68-73.
24. Sakaguchi, S., M. Ono, R. Setoguchi, et al., Foxp3+ CD25+ CD4+ natural regulatory T cells in dominant self-tolerance and autoimmune disease. Immunol Rev, 2006. 212: p. 8-27.
25. Ochs, H.D., E. Gambineri, and T.R. Torgerson, IPEX, FOXP3 and regulatory T-cells: a model for autoimmunity. Immunol Res, 2007. 38(1-3): p. 112-21.
26. Kim, J.M., J.P. Rasmussen, and A.Y. Rudensky, Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. Nat Immunol, 2007. 8(2): p. 191-7.
27. Sakaguchi, S., T. Yamaguchi, T. Nomura, et al., Regulatory T cells and immune tolerance. Cell, 2008. 133(5): p. 775-87.
28. Le Douarin, N., C. Corbel, A. Bandeira, et al., Evidence for a thymus-dependent form of tolerance that is not based on elimination or anergy of reactive T cells. Immunol Rev, 1996. 149: p. 35-53.
29. Cantor, H. and E. Simpson, Regulation of the immune response by subclasses of T lymphocytes. I. Interactions between pre-killer T cells and regulatory T cells obtained from peripheral lymphoid tissues of mice. Eur J Immunol, 1975. 5(5): p. 330-6.
30. Kindred, B. and B. Sordat, Lymphocytes which differentiate in an allogeneic thymus. II. Evidence for both central and peripheral mechanisms in tolerance to donor strain tissues. Eur J Immunol, 1977. 7(7): p. 437-42.
31. Scheinecker, C., M. Bonelli, and J.S. Smolen, Pathogenetic aspects of systemic lupus erythematosus with an emphasis on regulatory T cells. J Autoimmun, 2010. 35(3): p. 269-75.
32. Sakaguchi, S., K. Fukuma, K. Kuribayashi, et al., Organ-specific autoimmune diseases induced in mice by elimination of T cell subset. I. Evidence for the active participation of T cells in natural self-tolerance; deficit of a T cell subset as a possible cause of autoimmune disease. J Exp Med, 1985. 161(1): p. 72-87.
33. Sakaguchi, S., N. Sakaguchi, M. Asano, et al., Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. J Immunol, 1995. 155(3): p. 1151-64.
34. Suri-Payer, E., A.Z. Amar, A.M. Thornton, et al., CD4+CD25+ T cells inhibit both the induction and effector function of autoreactive T cells and represent a unique lineage of immunoregulatory cells. J Immunol, 1998. 160(3): p. 1212-8.
35. Levings, M.K., R. Sangregorio, and M.G. Roncarolo, Human cd25(+)cd4(+) t regulatory cells suppress naive and memory T cell proliferation and can be expanded in vitro without loss of function. J Exp Med, 2001. 193(11): p. 1295-302.
36. Baecher-Allan, C., J.A. Brown, G.J. Freeman, et al., CD4+CD25high regulatory cells in human peripheral blood. J Immunol, 2001. 167(3): p. 1245-53.
37. Fontenot, J.D., M.A. Gavin, and A.Y. Rudensky, Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. Nat Immunol, 2003. 4(4): p. 330-6.
38. Hori, S., T. Nomura, and S. Sakaguchi, Control of regulatory T cell development by the transcription factor Foxp3. Science, 2003. 299(5609): p. 1057-61.
39. Khattri, R., T. Cox, S.A. Yasayko, et al., An essential role for Scurfin in CD4+CD25+ T regulatory cells. Nat Immunol, 2003. 4(4): p. 337-42.
40. Seddiki, N., B. Santner-Nanan, J. Martinson, et al., Expression of interleukin (IL)-2 and IL-7 receptors discriminates between human regulatory and activated T cells. J Exp Med, 2006. 203(7): p. 1693-700.
41. Liu, W., A.L. Putnam, Z. Xu-Yu, et al., CD127 expression inversely correlates with FoxP3 and suppressive function of human CD4+ T reg cells. J Exp Med, 2006. 203(7): p. 1701-11.
42. Azuma, T., T. Takahashi, A. Kunisato, et al., Human CD4+ CD25+ regulatory T cells suppress NKT cell functions. Cancer Res, 2003. 63(15): p. 4516-20.
43. Frimpong-Boateng, K., N. van Rooijen, and R. Geiben-Lynn, Regulatory T cells

suppress natural killer cells during plasmid DNA vaccination in mice, blunting the CD8+ T cell immune response by the cytokine TGFbeta. PLoS One, 2010. 5(8): p. e12281.
44. Sakaguchi, S., M. Miyara, C.M. Costantino, et al., FOXP3+ regulatory T cells in the human immune system. Nat Rev Immunol, 2010. 10(7): p. 490-500.
45. Shevach, E.M., Mechanisms of foxp3+ T regulatory cell-mediated suppression. Immunity, 2009. 30(5): p. 636-45.
46. Samy, E.T., L.A. Parker, C.P. Sharp, et al., Continuous control of autoimmune disease by antigen-dependent polyclonal CD4+CD25+ regulatory T cells in the regional lymph node. J Exp Med, 2005. 202(6): p. 771-81.
47. Graca, L., A. Le Moine, C.Y. Lin, et al., Donor-specific transplantation tolerance: the paradoxical behavior of CD4+CD25+ T cells. Proc Natl Acad Sci U S A, 2004. 101(27): p. 10122-6.
48. Davies, J.D., L.Y. Leong, A. Mellor, et al., T cell suppression in transplantation tolerance through linked recognition. J Immunol, 1996. 156(10): p. 3602-7.
49. Flores-Borja, F., E.C. Jury, C. Mauri, et al., Defects in CTLA-4 are associated with abnormal regulatory T cell function in rheumatoid arthritis. Proc Natl Acad Sci U S A, 2008. 105(49): p. 19396-401.
50. Qureshi, O.S., Y. Zheng, K. Nakamura, et al., Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4. Science, 2011. 332(6029): p. 600-3.
51. Wing, K., T. Yamaguchi, and S. Sakaguchi, Cell-autonomous and -non-autonomous roles of CTLA-4 in immune regulation. Trends Immunol, 2011. 32(9): p. 428-33.
52. Deaglio, S. and S.C. Robson, Ectonucleotidases as regulators of purinergic signaling in thrombosis, inflammation, and immunity. Adv Pharmacol, 2011. 61: p. 301-32.
53. Sitkovsky, M., D. Lukashev, S. Deaglio, et al., Adenosine A2A receptor antagonists: blockade of adenosinergic effects and T regulatory cells. Br J Pharmacol, 2008. 153 Suppl 1: p. S457-64.
54. Whiteside, T.L., M. Mandapathil, and P. Schuler, The role of the adenosinergic pathway in immunosuppression mediated by human regulatory T cells (treg). Curr Med Chem, 2011. 18(34): p. 5217-23.
55. Kobie, J.J., P.R. Shah, L. Yang, et al., T regulatory and primed uncommitted CD4 T cells express CD73, which suppresses effector CD4 T cells by converting 5'-adenosine monophosphate to adenosine. J Immunol, 2006. 177(10): p. 6780-6.
56. Wohler, J., D. Bullard, T. Schoeb, et al., LFA-1 is critical for regulatory T cell homeostasis and function. Mol Immunol, 2009. 46(11-12): p. 2424-8.
57. Huang, C.T., C.J. Workman, D. Flies, et al., Role of LAG-3 in regulatory T cells. Immunity, 2004. 21(4): p. 503-13.
58. Liao, G., S. Nayak, J.R. Regueiro, et al., GITR engagement preferentially enhances proliferation of functionally competent CD4+CD25+FoxP3+ regulatory T cells. Int Immunol, 2010. 22(4): p. 259-70.
59. Shevach, E.M., R.A. DiPaolo, J. Andersson, et al., The lifestyle of naturally occurring CD4+ CD25+ Foxp3+ regulatory T cells. Immunol Rev, 2006. 212: p. 60-73.
60. Cao, X., S.F. Cai, T.A. Fehniger, et al., Granzyme B and perforin are important for regulatory T cell-mediated suppression of tumor clearance. Immunity, 2007. 27(4): p. 635-46.
61. Grossman, W.J., J.W. Verbsky, W. Barchet, et al., Human T regulatory cells can use the perforin pathway to cause autologous target cell death. Immunity, 2004. 21(4): p. 589-601.
62. Mercer, F., L. Kozhaya, and D. Unutmaz, Expression and function of TNF and IL-1 receptors on human regulatory T cells. PLoS One, 2010. 5(1): p. e8639.
63. Tran, D.Q., TGF-beta: the sword, the wand, and the shield of FOXP3+ regulatory T cells. J Mol Cell Biol, 2011.
64. Shen, E., K. Zhao, C. Wu, et al., The suppressive effect of CD25+Treg cells on Th1 differentiation requires cell-cell contact partially via TGF-beta production. Cell Biol Int, 2011. 35(7): p. 705-12.
65. Chen, M.L., M.J. Pittet, L. Gorelik, et al., Regulatory T cells suppress tumor-specific CD8 T cell cytotoxicity through TGF-beta signals in vivo. Proc Natl Acad Sci U S A, 2005. 102(2): p. 419-24.
66. Fujio, K., T. Okamura, and K. Yamamoto, The Family of IL-10-secreting CD4+ T

cells. Adv Immunol, 2010. 105: p. 99-130.
67. McGeachy, M.J., L.A. Stephens, and S.M. Anderton, Natural recovery and protection from autoimmune encephalomyelitis: contribution of CD4+CD25+ regulatory cells within the central nervous system. J Immunol, 2005. 175(5): p. 3025-32.
68. Bettini, M. and D.A. Vignali, Regulatory T cells and inhibitory cytokines in autoimmunity. Curr Opin Immunol, 2009. 21(6): p. 612-8.
69. Selvaraj, R.K. and T.L. Geiger, Mitigation of experimental allergic encephalomyelitis by TGF-beta induced Foxp3+ regulatory T lymphocytes through the induction of anergy and infectious tolerance. J Immunol, 2008. 180(5): p. 2830-8.
70. Ghiringhelli, F., C. Menard, F. Martin, et al., The role of regulatory T cells in the control of natural killer cells: relevance during tumor progression. Immunol Rev, 2006. 214: p. 229-38.
71. Collison, L.W., C.J. Workman, T.T. Kuo, et al., The inhibitory cytokine IL-35 contributes to regulatory T-cell function. Nature, 2007. 450(7169): p. 566-9.
72. Bardel, E., F. Larousserie, P. Charlot-Rabiega, et al., Human CD4+ CD25+ Foxp3+ regulatory T cells do not constitutively express IL-35. J Immunol, 2008. 181(10): p. 6898-905.
73. Vokaer, B., N. Van Rompaey, P.H. Lemaitre, et al., Critical role of regulatory T cells in Th17-mediated minor antigen-disparate rejection. J Immunol, 2010. 185(6): p. 3417-25.
74. Allan, S.E., R. Broady, S. Gregori, et al., CD4+ T-regulatory cells: toward therapy for human diseases. Immunol Rev, 2008. 223: p. 391-421.
75. Wing, K., A. Ekmark, H. Karlsson, et al., Characterization of human CD25+ CD4+ T cells in thymus, cord and adult blood. Immunology, 2002. 106(2): p. 190-9.
76. Chatenoud, L., J. Primo, and J.F. Bach, CD3 antibody-induced dominant self tolerance in overtly diabetic NOD mice. J Immunol, 1997. 158(6): p. 2947-54.
77. Zheng, S.G., J.H. Wang, J.D. Gray, et al., Natural and induced CD4+CD25+ cells educate CD4+CD25- cells to develop suppressive activity: the role of IL-2, TGF-beta, and IL-10. J Immunol, 2004. 172(9): p. 5213-21.
78. Zheng, S.G., J. Wang, P. Wang, et al., IL-2 is essential for TGF-beta to convert naive CD4+CD25- cells to CD25+Foxp3+ regulatory T cells and for expansion of these cells. J Immunol, 2007. 178(4): p. 2018-27.
79. Andersson, J., D.Q. Tran, M. Pesu, et al., CD4+ FoxP3+ regulatory T cells confer infectious tolerance in a TGF-beta-dependent manner. J Exp Med, 2008. 205(9): p. 1975-81.
80. Vukmanovic-Stejic, M., Y. Zhang, J.E. Cook, et al., Human CD4+ CD25hi Foxp3+ regulatory T cells are derived by rapid turnover of memory populations in vivo. J Clin Invest, 2006. 116(9): p. 2423-33.
81. Horwitz, D.A., S.G. Zheng, and J.D. Gray, Natural and TGF-beta-induced Foxp3(+)CD4(+) CD25(+) regulatory T cells are not mirror images of each other. Trends Immunol, 2008. 29(9): p. 429-35.
82. Curotto de Lafaille, M.A. and J.J. Lafaille, Natural and adaptive foxp3+ regulatory T cells: more of the same or a division of labor? Immunity, 2009. 30(5): p. 626-35.
83. Bluestone, J.A. and A.K. Abbas, Natural versus adaptive regulatory T cells. Nat Rev Immunol, 2003. 3(3): p. 253-7.
84. Coombes, J.L., K.R. Siddiqui, C.V. Arancibia-Carcamo, et al., A functionally specialized population of mucosal CD103+ DCs induces Foxp3+ regulatory T cells via a TGF-beta and retinoic acid-dependent mechanism. J Exp Med, 2007. 204(8): p. 1757-64.
85. Zheng, Y., S. Josefowicz, A. Chaudhry, et al., Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. Nature, 2010. 463(7282): p. 808-12.
86. Gavin, M.A., T.R. Torgerson, E. Houston, et al., Single-cell analysis of normal and FOXP3-mutant human T cells: FOXP3 expression without regulatory T cell development. Proc Natl Acad Sci U S A, 2006. 103(17): p. 6659-64.
87. Lu, L., X. Zhou, J. Wang, et al., Characterization of protective human CD4CD25 FOXP3 regulatory T cells generated with IL-2, TGF-beta and retinoic acid. PLoS One, 2010. 5(12): p. e15150.
88. Gandhi, R., D. Kumar, E.J. Burns, et al., Activation of the aryl hydrocarbon receptor induces human type 1 regulatory T cell-like and Foxp3(+) regulatory T cells. Nat Immunol, 2010. 11(9): p. 846-53.

89. Thornton, A.M., P.E. Korty, D.Q. Tran, et al., Expression of Helios, an Ikaros transcription factor family member, differentiates thymic-derived from peripherally induced Foxp3+ T regulatory cells. J Immunol, 2010. 184(7): p. 3433-41.
90. Akimova, T., U.H. Beier, L. Wang, et al., Helios expression is a marker of T cell activation and proliferation. PLoS One, 2011. 6(8): p. e24226.
91. Gottschalk, R.A., E. Corse, and J.P. Allison, Expression of helios in peripherally induced foxp3+ regulatory T cells. J Immunol, 2012. 188(3): p. 976-80.
92. Serre, K., C. Benezech, G. Desanti, et al., Helios is associated with CD4 T cells differentiating to T helper 2 and follicular helper T cells in vivo independently of Foxp3 expression. PLoS One, 2011. 6(6): p. e20731.
93. Getnet, D., J.F. Grosso, M.V. Goldberg, et al., A role for the transcription factor Helios in human CD4(+)CD25(+) regulatory T cells. Mol Immunol, 2010. 47(7-8): p. 1595-600.
94. Schmidl, C., L. Hansmann, R. Andreesen, et al., Epigenetic reprogramming of the RORC locus during in vitro expansion is a distinctive feature of human memory but not naive Treg. Eur J Immunol, 2011. 41(5): p. 1491-8.
95. Ayyoub, M., F. Deknuydt, I. Raimbaud, et al., Human memory FOXP3+ Tregs secrete IL-17 ex vivo and constitutively express the T(H)17 lineage-specific transcription factor RORgamma t. Proc Natl Acad Sci U S A, 2009. 106(21): p. 8635-40.
96. Miyara, M., Y. Yoshioka, A. Kitoh, et al., Functional delineation and differentiation dynamics of human CD4+ T cells expressing the FoxP3 transcription factor. Immunity, 2009. 30(6): p. 899-911.
97. Tang, Q., K.J. Henriksen, M. Bi, et al., In vitro-expanded antigen-specific regulatory T cells suppress autoimmune diabetes. J Exp Med, 2004. 199(11): p. 1455-65.
98. Kohm, A.P., P.A. Carpentier, H.A. Anger, et al., Cutting edge: CD4+CD25+ regulatory T cells suppress antigen-specific autoreactive immune responses and central nervous system inflammation during active experimental autoimmune encephalomyelitis. J Immunol, 2002. 169(9): p. 4712-6.
99. Haas, J., A. Hug, A. Viehover, et al., Reduced suppressive effect of CD4+CD25high regulatory T cells on the T cell immune response against myelin oligodendrocyte glycoprotein in patients with multiple sclerosis. Eur J Immunol, 2005. 35(11): p. 3343-52.
100. Long, S.A., K. Cerosaletti, P.L. Bollyky, et al., Defects in IL-2R signaling contribute to diminished maintenance of FOXP3 expression in CD4(+)CD25(+) regulatory T-cells of type 1 diabetic subjects. Diabetes, 2010. 59(2): p. 407-15.
101. Willcox, A., S.J. Richardson, A.J. Bone, et al., Analysis of islet inflammation in human type 1 diabetes. Clin Exp Immunol, 2009. 155(2): p. 173-81.
102. Buckner, J.H., Mechanisms of impaired regulation by CD4(+)CD25(+)FOXP3(+) regulatory T cells in human autoimmune diseases. Nat Rev Immunol, 2010. 10(12): p. 849-59.
103. Kumar, M., N. Putzki, V. Limmroth, et al., CD4+CD25+FoxP3+ T lymphocytes fail to suppress myelin basic protein-induced proliferation in patients with multiple sclerosis. J Neuroimmunol, 2006. 180(1-2): p. 178-84.
104. Feger, U., C. Luther, S. Poeschel, et al., Increased frequency of CD4+ CD25+ regulatory T cells in the cerebrospinal fluid but not in the blood of multiple sclerosis patients. Clin Exp Immunol, 2007. 147(3): p. 412-8.
105. de Andres, C., C. Aristimuno, V. de Las Heras, et al., Interferon beta-1a therapy enhances CD4+ regulatory T-cell function: an ex vivo and in vitro longitudinal study in relapsing-remitting multiple sclerosis. J Neuroimmunol, 2007. 182(1-2): p. 204-11.
106. Horwitz, D.A., Identity of mysterious CD4+CD25-Foxp3+ cells in SLE. Arthritis Res Ther, 2010. 12(1): p. 101.
107. Isenberg, D.A., J.J. Manson, M.R. Ehrenstein, et al., Fifty years of anti-ds DNA antibodies: are we approaching journey's end? Rheumatology (Oxford), 2007. 46(7): p. 1052-6.
108. Rahman, A. and D.A. Isenberg, Systemic lupus erythematosus. N Engl J Med, 2008. 358(9): p. 929-39.
109. .
110. National Institute of Arthritis and Musculoskeletal ans Skin Diseases (NIAMS) Image Gallery. Available from: http://images.niams.nih.gov/index.cfm (Accessed 12 Sept 2014).

111. Schur P.H., H., B. H. Epidemiology and pathogenesis of systemic lupus erythematosus. 2014; Available from: http://www.uptodate.com/contents/epidemiology-and-pathogenesis-of-systemic-lupus-erythematosus (Accessed Sept 2014).
112. Simard, J.F. and K.H. Costenbader, What can epidemiology tell us about systemic lupus erythematosus? Int J Clin Pract, 2007. 61(7): p. 1170-80.
113. Bernatsky, S., J.F. Boivin, L. Joseph, et al., Mortality in systemic lupus erythematosus. Arthritis Rheum, 2006. 54(8): p. 2550-7.
114. Cervera, R., M.A. Khamashta, J. Font, et al., Morbidity and mortality in systemic lupus erythematosus during a 10-year period: a comparison of early and late manifestations in a cohort of 1,000 patients. Medicine (Baltimore), 2003. 82(5): p. 299-308.
115. Fairweather, D., S. Frisancho-Kiss, and N.R. Rose, Sex differences in autoimmune disease from a pathological perspective. Am J Pathol, 2008. 173(3): p. 600-9.
116. Samstein, R.M., S.Z. Josefowicz, A. Arvey, et al., Extrathymic generation of regulatory T cells in placental mammals mitigates maternal-fetal conflict. Cell, 2012. 150(1): p. 29-38.
117. Block, S.R., J.B. Winfield, M.D. Lockshin, et al., Studies of twins with systemic lupus erythematosus. A review of the literature and presentation of 12 additional sets. Am J Med, 1975. 59(4): p. 533-52.
118. Deapen, D., A. Escalante, L. Weinrib, et al., A revised estimate of twin concordance in systemic lupus erythematosus. Arthritis Rheum, 1992. 35(3): p. 311-8.
119. Malik, U.R., D.F. Makower, and S. Wadler, Interferon-mediated fatigue. Cancer, 2001. 92(6 Suppl): p. 1664-8.
120. Ng, W.F. and S.J. Bowman, Primary Sjogren's syndrome: too dry and too tired. Rheumatology (Oxford), 2010. 49(5): p. 844-53.
121. Fonseca, R., M. Bernardes, G. Terroso, et al., Silent Burdens in Disease: Fatigue and Depression in SLE. Autoimmune Dis, 2014. 2014: p. 790724.
122. Systemic lupus erythematosus (autoantibody-positive) - belimumab [ID416]. Available from: https://www.nice.org.uk/Guidance/InDevelopment/GID-TAG273 (Accessed Sept 2014).
123. Huscher, D., K. Thiele, E. Gromnica-Ihle, et al., Dose-related patterns of glucocorticoid-induced side effects. Ann Rheum Dis, 2009. 68(7): p. 1119-24.
124. Lo, M.S. and G.C. Tsokos, Treatment of systemic lupus erythematosus: new advances in targeted therapy. Ann N Y Acad Sci, 2012.
125. Stohl, W., Future prospects in biologic therapy for systemic lupus erythematosus. Nat Rev Rheumatol, 2013. 9(12): p. 705-20.
126. Leandro, M.J., J.C. Edwards, G. Cambridge, et al., An open study of B lymphocyte depletion in systemic lupus erythematosus. Arthritis Rheum, 2002. 46(10): p. 2673-7.
127. Looney, R.J., J.H. Anolik, D. Campbell, et al., B cell depletion as a novel treatment for systemic lupus erythematosus: a phase I/II dose-escalation trial of rituximab. Arthritis Rheum, 2004. 50(8): p. 2580-9.
128. Merrill, J.T., C.M. Neuwelt, D.J. Wallace, et al., Efficacy and safety of rituximab in moderately-to-severely active systemic lupus erythematosus: the randomized, double-blind, phase II/III systemic lupus erythematosus evaluation of rituximab trial. Arthritis Rheum, 2010. 62(1): p. 222-33.
129. Rovin, B.H., R. Furie, K. Latinis, et al., Efficacy and safety of rituximab in patients with active proliferative lupus nephritis: the Lupus Nephritis Assessment with Rituximab study. Arthritis Rheum, 2012. 64(4): p. 1215-26.
130. Stohl, W., F. Hiepe, K.M. Latinis, et al., Belimumab reduces autoantibodies, normalizes low complement, and reduces select B-cell populations in patients with systemic lupus erythematosus. Arthritis Rheum, 2012.
131. EMA, Assessment report: Benlysta.
132. Illei, G.G., R. Cervera, R.K. Burt, et al., Current state and future directions of autologous hematopoietic stem cell transplantation in systemic lupus erythematosus. Ann Rheum Dis, 2011. 70(12): p. 2071-4.
133. Marmont du Haut Champ, A.M., Hematopoietic stem cell transplantation for systemic lupus erythematosus. Clin Dev Immunol, 2012. 2012: p. 380391.
134. Dorner, T., A.M. Jacobi, and P.E. Lipsky, B cells in autoimmunity. Arthritis Res Ther, 2009. 11(5): p. 247.

135. Munoz, L.E., K. Lauber, M. Schiller, et al., The role of defective clearance of apoptotic cells in systemic autoimmunity. Nat Rev Rheumatol, 2010. 6(5): p. 280-9.
136. Baumann, I., W. Kolowos, R.E. Voll, et al., Impaired uptake of apoptotic cells into tingible body macrophages in germinal centers of patients with systemic lupus erythematosus. Arthritis Rheum, 2002. 46(1): p. 191-201.
137. Elkon, K.B. and D.M. Santer, Complement, interferon and lupus. Curr Opin Immunol, 2012. 24(6): p. 665-70.
138. Liu, B., Y. Yang, J. Dai, et al., TLR4 up-regulation at protein or gene level is pathogenic for lupus-like autoimmune disease. J Immunol, 2006. 177(10): p. 6880-8.
139. Hackl, D., J. Loschko, T. Sparwasser, et al., Activation of dendritic cells via TLR7 reduces Foxp3 expression and suppressive function in induced Tregs. Eur J Immunol, 2011. 41(5): p. 1334-43.
140. Christensen, S.R., J. Shupe, K. Nickerson, et al., Toll-like receptor 7 and TLR9 dictate autoantibody specificity and have opposing inflammatory and regulatory roles in a murine model of lupus. Immunity, 2006. 25(3): p. 417-28.
141. Guggino, G., A.R. Giardina, F. Ciccia, et al., Are Toll-like receptors and decoy receptors involved in the immunopathogenesis of systemic lupus erythematosus and lupus-like syndromes? Clin Dev Immunol, 2012. 2012: p. 135932.
142. Karageorgas, T.P., D.D. Tseronis, and C.P. Mavragani, Activation of type I interferon pathway in systemic lupus erythematosus: association with distinct clinical phenotypes. J Biomed Biotechnol, 2011. 2011: p. 273907.
143. Bengtsson, A.A., G. Sturfelt, L. Truedsson, et al., Activation of type I interferon system in systemic lupus erythematosus correlates with disease activity but not with antiretroviral antibodies. Lupus, 2000. 9(9): p. 664-71.
144. Bennett, L., A.K. Palucka, E. Arce, et al., Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med, 2003. 197(6): p. 711-23.
145. Kirou, K.A., C. Lee, S. George, et al., Coordinate overexpression of interferon-alpha-induced genes in systemic lupus erythematosus. Arthritis Rheum, 2004. 50(12): p. 3958-67.
146. Chaussabel, D., C. Quinn, J. Shen, et al., A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity, 2008. 29(1): p. 150-64.
147. Fitzgerald-Bocarsly, P., J. Dai, and S. Singh, Plasmacytoid dendritic cells and type I IFN: 50 years of convergent history. Cytokine Growth Factor Rev, 2008. 19(1): p. 3-19.
148. Niewold, T.B., D.N. Clark, R. Salloum, et al., Interferon alpha in systemic lupus erythematosus. J Biomed Biotechnol, 2010. 2010: p. 948364.
149. Aringer, M., C. Gunther, and M.A. Lee-Kirsch, Innate immune processes in lupus erythematosus. Clin Immunol, 2013. 147(3): p. 216-22.
150. Studnicka-Benke, A., G. Steiner, P. Petera, et al., Tumour necrosis factor alpha and its soluble receptors parallel clinical disease and autoimmune activity in systemic lupus erythematosus. Br J Rheumatol, 1996. 35(11): p. 1067-74.
151. Ohl, K. and K. Tenbrock, Inflammatory cytokines in systemic lupus erythematosus. J Biomed Biotechnol, 2011. 2011: p. 432595.
152. Wong, C.K., L.C. Lit, L.S. Tam, et al., Hyperproduction of IL-23 and IL-17 in patients with systemic lupus erythematosus: implications for Th17-mediated inflammation in auto-immunity. Clin Immunol, 2008. 127(3): p. 385-93.
153. Aringer, M., G.H. Stummvoll, G. Steiner, et al., Serum interleukin-15 is elevated in systemic lupus erythematosus. Rheumatology (Oxford), 2001. 40(8): p. 876-81.
154. Crispin, J.C., M. Oukka, G. Bayliss, et al., Expanded double negative T cells in patients with systemic lupus erythematosus produce IL-17 and infiltrate the kidneys. J Immunol, 2008. 181(12): p. 8761-6.
155. Alcocer-Varela, J. and D. Alarcon-Segovia, Decreased production of and response to interleukin-2 by cultured lymphocytes from patients with systemic lupus erythematosus. J Clin Invest, 1982. 69(6): p. 1388-92.
156. Kow, N.Y. and A. Mak, Costimulatory pathways: physiology and potential therapeutic manipulation in systemic lupus erythematosus. Clin Dev Immunol, 2013. 2013: p. 245928.
157. Jury, E.C., P.S. Kabouridis, F. Flores-Borja, et al., Altered lipid raft-associated signaling and ganglioside expression in T lymphocytes from patients with systemic lupus erythematosus. J Clin Invest, 2004. 113(8): p. 1176-87.

158. Lyons, P.A., E.F. McKinney, T.F. Rayner, et al., Novel expression signatures identified by transcriptional analysis of separated leucocyte subsets in systemic lupus erythematosus and vasculitis. Ann Rheum Dis, 2010. 69(6): p. 1208-13.
159. Frangou, E.A., G.K. Bertsias, and D.T. Boumpas, Gene expression and regulation in systemic lupus erythematosus. Eur J Clin Invest, 2013. 43(10): p. 1084-96.
160. Hu, N., X. Qiu, Y. Luo, et al., Abnormal histone modification patterns in lupus CD4+ T cells. J Rheumatol, 2008. 35(5): p. 804-10.
161. Yan, K., Q. Cao, C.M. Reilly, et al., Histone deacetylase 9 deficiency protects against effector T cell-mediated systemic autoimmunity. J Biol Chem, 2011. 286(33): p. 28833-43.
162. Zhang, Q., H. Long, J. Liao, et al., Inhibited expression of hematopoietic progenitor kinase 1 associated with loss of jumonji domain containing 3 promoter binding contributes to autoimmunity in systemic lupus erythematosus. J Autoimmun, 2011. 37(3): p. 180-9.
163. Huang, X., Y. Guo, C. Bao, et al., Multidimensional single cell based STAT phosphorylation profiling identifies a novel biosignature for evaluation of systemic lupus erythematosus activity. PLoS One, 2011. 6(7): p. e21671.
164. Liang, Y., W.D. Xu, X.K. Yang, et al., Association of signaling transducers and activators of transcription 1 and systemic lupus erythematosus. Autoimmunity, 2014. 47(3): p. 141-5.
165. Miyara, M., G. Gorochov, M. Ehrenstein, et al., Human FoxP3+ regulatory T cells in systemic autoimmune diseases. Autoimmun Rev, 2011. 10(12): p. 744-55.
166. Lyssuk, E.Y., A.V. Torgashina, S.K. Soloviev, et al., Reduced number and function of CD4+CD25highFoxP3+ regulatory T cells in patients with systemic lupus erythematosus. Adv Exp Med Biol, 2007. 601: p. 113-9.
167. Barreto, M., R.C. Ferreira, L. Lourenco, et al., Low frequency of CD4+CD25+ Treg in SLE patients: a heritable trait associated with CTLA4 and TGFbeta gene variants. BMC Immunol, 2009. 10: p. 5.
168. Fathy, A., R.W. Mohamed, G.A. Tawfik, et al., Diminished CD4+CD25+ T-lymphocytes in peripheral blood of patients with systemic lupus erythematosus. Egypt J Immunol, 2005. 12(1): p. 25-31.
169. Lee, J.H., L.C. Wang, Y.T. Lin, et al., Inverse correlation between CD4+ regulatory T-cell population and autoantibody levels in paediatric patients with systemic lupus erythematosus. Immunology, 2006. 117(2): p. 280-6.
170. Azab, N.A., I.H. Bassyouni, Y. Emad, et al., CD4+CD25+ regulatory T cells (TREG) in systemic lupus erythematosus (SLE) patients: the possible influence of treatment with corticosteroids. Clin Immunol, 2008. 127(2): p. 151-7.
171. Crispin, J.C., A. Martinez, and J. Alcocer-Varela, Quantification of regulatory T cells in patients with systemic lupus erythematosus. J Autoimmun, 2003. 21(3): p. 273-6.
172. Miyara, M., Z. Amoura, C. Parizot, et al., Global natural regulatory T cell depletion in active systemic lupus erythematosus. J Immunol, 2005. 175(12): p. 8392-400.
173. Mellor-Pita, S., M.J. Citores, R. Castejon, et al., Decrease of regulatory T cells in patients with systemic lupus erythematosus. Ann Rheum Dis, 2006. 65(4): p. 553-4.
174. Bonelli, M., K. von Dalwigk, A. Savitskaya, et al., Foxp3 expression in CD4+ T cells of patients with systemic lupus erythematosus: a comparative phenotypic analysis. Ann Rheum Dis, 2008. 67(5): p. 664-71.
175. Lee, H.Y., Y.K. Hong, H.J. Yun, et al., Altered frequency and migration capacity of CD4+CD25+ regulatory T cells in systemic lupus erythematosus. Rheumatology (Oxford), 2008. 47(6): p. 789-94.
176. Bonelli, M., A. Savitskaya, K. von Dalwigk, et al., Quantitative and qualitative deficiencies of regulatory T cells in patients with systemic lupus erythematosus (SLE). Int Immunol, 2008. 20(7): p. 861-8.
177. Alvarado-Sanchez, B., B. Hernandez-Castro, D. Portales-Perez, et al., Regulatory T cells in patients with systemic lupus erythematosus. J Autoimmun, 2006. 27(2): p. 110-8.
178. Yates, J., A. Whittington, P. Mitchell, et al., Natural regulatory T cells: number and function are normal in the majority of patients with lupus nephritis. Clin Exp Immunol, 2008. 153(1): p. 44-55.
179. Suarez, A., P. Lopez, J. Gomez, et al., Enrichment of CD4+ CD25high T cell population in patients with systemic lupus erythematosus treated with glucocorticoids. Ann Rheum Dis, 2006. 65(11): p. 1512-7.

180. Yang, J., Y. Chu, X. Yang, et al., Th17 and natural Treg cell population dynamics in systemic lupus erythematosus. Arthritis Rheum, 2009. 60(5): p. 1472-83.
181. Henriques, A., L. Ines, M. Couto, et al., Frequency and functional activity of Th17, Tc17 and other T-cell subsets in Systemic Lupus Erythematosus. Cell Immunol, 2010. 264(1): p. 97-103.
182. Jury, E.C., F. Flores-Borja, H.S. Kalsi, et al., Abnormal CTLA-4 function in T cells from patients with systemic lupus erythematosus. Eur J Immunol, 2010. 40(2): p. 569-78.
183. Zhang, B., X. Zhang, F. Tang, et al., Reduction of forkhead box P3 levels in CD4+CD25high T cells in patients with new-onset systemic lupus erythematosus. Clin Exp Immunol, 2008. 153(2): p. 182-7.
184. Lin, S.C., K.H. Chen, C.H. Lin, et al., The quantitative analysis of peripheral blood FOXP3-expressing T cells in systemic lupus erythematosus and rheumatoid arthritis patients. Eur J Clin Invest, 2007. 37(12): p. 987-96.
185. Vargas-Rojas, M.I., J.C. Crispin, Y. Richaud-Patin, et al., Quantitative and qualitative normal regulatory T cells are not capable of inducing suppression in SLE patients due to T-cell resistance. Lupus, 2008. 17(4): p. 289-94.
186. Yan, B., S. Ye, G. Chen, et al., Dysfunctional CD4+,CD25+ regulatory T cells in untreated active systemic lupus erythematosus secondary to interferon-alpha-producing antigen-presenting cells. Arthritis Rheum, 2008. 58(3): p. 801-12.
187. Barath, S., M. Aleksza, T. Tarr, et al., Measurement of natural (CD4+CD25high) and inducible (CD4+IL-10+) regulatory T cells in patients with systemic lupus erythematosus. Lupus, 2007. 16(7): p. 489-96.
188. Suen, J.L., H.T. Li, Y.J. Jong, et al., Altered homeostasis of CD4(+) FoxP3(+) regulatory T-cell subpopulations in systemic lupus erythematosus. Immunology, 2009. 127(2): p. 196-205.
189. Venigalla, R.K., T. Tretter, S. Krienke, et al., Reduced CD4+,CD25- T cell sensitivity to the suppressive function of CD4+,CD25high,CD127 -/low regulatory T cells in patients with active systemic lupus erythematosus. Arthritis Rheum, 2008. 58(7): p. 2120-30.
190. Fontenot, J.D., J.P. Rasmussen, M.A. Gavin, et al., A function for interleukin 2 in Foxp3-expressing regulatory T cells. Nat Immunol, 2005. 6(11): p. 1142-51.
191. Setoguchi, R., S. Hori, T. Takahashi, et al., Homeostatic maintenance of natural Foxp3(+) CD25(+) CD4(+) regulatory T cells by interleukin (IL)-2 and induction of autoimmune disease by IL-2 neutralization. J Exp Med, 2005. 201(5): p. 723-35.
192. Solomou, E.E., Y.T. Juang, M.F. Gourley, et al., Molecular basis of deficient IL-2 production in T cells from patients with systemic lupus erythematosus. J Immunol, 2001. 166(6): p. 4216-22.
193. Valencia, X., C. Yarboro, G. Illei, et al., Deficient CD4+CD25high T regulatory cell function in patients with active systemic lupus erythematosus. J Immunol, 2007. 178(4): p. 2579-88.
194. Walker, M.R., D.J. Kasprowicz, V.H. Gersuk, et al., Induction of FoxP3 and acquisition of T regulatory activity by stimulated human CD4+CD25- T cells. J Clin Invest, 2003. 112(9): p. 1437-43.
195. Xu, L., A. Kitani, I. Fuss, et al., Cutting edge: regulatory T cells induce CD4+CD25-Foxp3- T cells or are self-induced to become Th17 cells in the absence of exogenous TGF-beta. J Immunol, 2007. 178(11): p. 6725-9.
196. Ouyang, W., J.K. Kolls, and Y. Zheng, The biological functions of T helper 17 cell effector cytokines in inflammation. Immunity, 2008. 28(4): p. 454-67.
197. Valencia, X., G. Stephens, R. Goldbach-Mansky, et al., TNF downmodulates the function of human CD4+CD25hi T-regulatory cells. Blood, 2006. 108(1): p. 253-61.
198. Lin, Y.C., J.H. Lee, A.S. Wu, et al., Association of single-nucleotide polymorphisms in FOXP3 gene with systemic lupus erythematosus susceptibility: a case-control study. Lupus, 2011. 20(2): p. 137-43.
199. Brownlie, R.J. and R. Zamoyska, T cell receptor signalling networks: branched, diversified and bounded. Nat Rev Immunol, 2013. 13(4): p. 257-69.
200. Chen, L. and D.B. Flies, Molecular mechanisms of T cell co-stimulation and co-inhibition. Nat Rev Immunol, 2013. 13(4): p. 227-42.
201. Zeiser, R., D.B. Leveson-Gower, E.A. Zambricki, et al., Differential impact of mammalian target of rapamycin inhibition on CD4+CD25+Foxp3+ regulatory T cells compared with conventional CD4+ T cells. Blood, 2008. 111(1): p. 453-62.
202. Battaglia, M., A. Stabilini, and M.G. Roncarolo, Rapamycin selectively expands

CD4+CD25+FoxP3+ regulatory T cells. Blood, 2005. 105(12): p. 4743-8.
203. Taguchi, O., K. Kontani, H. Ikeda, et al., Tissue-specific suppressor T cells involved in self-tolerance are activated extrathymically by self-antigens. Immunology, 1994. 82(3): p. 365-9.
204. Jordan, M.S., A. Boesteanu, A.J. Reed, et al., Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. Nat Immunol, 2001. 2(4): p. 301-6.
205. Maloy, K.J. and F. Powrie, Regulatory T cells in the control of immune pathology. Nat Immunol, 2001. 2(9): p. 816-22.
206. Lee, H.M., J.L. Bautista, and C.S. Hsieh, Thymic and peripheral differentiation of regulatory T cells. Adv Immunol, 2011. 112: p. 25-71.
207. Papiernik, M., M.L. de Moraes, C. Pontoux, et al., Regulatory CD4 T cells: expression of IL-2R alpha chain, resistance to clonal deletion and IL-2 dependency. Int Immunol, 1998. 10(4): p. 371-8.
208. Furtado, G.C., M.A. Curotto de Lafaille, N. Kutchukhidze, et al., Interleukin 2 signaling is required for CD4(+) regulatory T cell function. J Exp Med, 2002. 196(6): p. 851-7.
209. Miyazaki, T., A. Kawahara, H. Fujii, et al., Functional activation of Jak1 and Jak3 by selective association with IL-2 receptor subunits. Science, 1994. 266(5187): p. 1045-7.
210. Lin, J.X. and W.J. Leonard, The role of Stat5a and Stat5b in signaling by IL-2 family cytokines. Oncogene, 2000. 19(21): p. 2566-76.
211. Zhang, Y., X. Feng, R. We, et al., Receptor-associated Mad homologues synergize as effectors of the TGF-beta response. Nature, 1996. 383(6596): p. 168-72.
212. Huse, M., T.W. Muir, L. Xu, et al., The TGF beta receptor activation process: an inhibitor- to substrate-binding switch. Mol Cell, 2001. 8(3): p. 671-82.
213. Jana, S., P. Jailwala, D. Haribhai, et al., The role of NF-kappaB and Smad3 in TGF-beta-mediated Foxp3 expression. Eur J Immunol, 2009. 39(9): p. 2571-83.
214. Chen, W. and J.E. Konkel, TGF-beta and 'adaptive' Foxp3(+) regulatory T cells. J Mol Cell Biol, 2010. 2(1): p. 30-6.
215. Yamaguchi, K., S. Nagai, J. Ninomiya-Tsuji, et al., XIAP, a cellular member of the inhibitor of apoptosis protein family, links the receptors to TAB1-TAK1 in the BMP signaling pathway. EMBO J, 1999. 18(1): p. 179-87.
216. Derynck, R. and Y.E. Zhang, Smad-dependent and Smad-independent pathways in TGF-beta family signalling. Nature, 2003. 425(6958): p. 577-84.
217. Chen, X., M. Baumel, D.N. Mannel, et al., Interaction of TNF with TNF receptor type 2 promotes expansion and function of mouse CD4+CD25+ T regulatory cells. J Immunol, 2007. 179(1): p. 154-61.
218. Chen, X., X. Wu, Q. Zhou, et al., TNFR2 is critical for the stabilization of the CD4+Foxp3+ regulatory T. cell phenotype in the inflammatory environment. J Immunol, 2013. 190(3): p. 1076-84.
219. Grinberg-Bleyer, Y., D. Saadoun, A. Baeyens, et al., Pathogenic T cells have a paradoxical protective effect in murine autoimmune diabetes by boosting Tregs. J Clin Invest, 2010. 120(12): p. 4558-68.
220. Chen, X., J.J. Subleski, R. Hamano, et al., Co-expression of TNFR2 and CD25 identifies more of the functional CD4+FOXP3+ regulatory T cells in human peripheral blood. Eur J Immunol, 2010. 40(4): p. 1099-106.
221. Hamano, R., J. Huang, T. Yoshimura, et al., TNF optimally activatives regulatory T cells by inducing TNF receptor superfamily members TNFR2, 4-1BB and OX40. Eur J Immunol, 2011. 41(7): p. 2010-20.
222. Zheng, W. and R.A. Flavell, The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. Cell, 1997. 89(4): p. 587-96.
223. Wei, G., B.J. Abraham, R. Yagi, et al., Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. Immunity, 2011. 35(2): p. 299-311.
224. Szabo, S.J., S.T. Kim, G.L. Costa, et al., A novel transcription factor, T-bet, directs Th1 lineage commitment. Cell, 2000. 100(6): p. 655-69.
225. Wilson, V. and F.L. Conlon, The T-box family. Genome Biol, 2002. 3(6): p. REVIEWS3008.
226. Ravindran, R., J. Foley, T. Stoklasek, et al., Expression of T-bet by CD4 T cells is essential for resistance to Salmonella infection. J Immunol, 2005. 175(7): p. 4603-

10.
227. Miller, S.A., A.C. Huang, M.M. Miazgowicz, et al., Coordinated but physically separable interaction with H3K27-demethylase and H3K4-methyltransferase activities are required for T-box protein-mediated activation of developmental gene expression. Genes Dev, 2008. 22(21): p. 2980-93.
228. Avni, O., D. Lee, F. Macian, et al., T(H) cell differentiation is accompanied by dynamic changes in histone acetylation of cytokine genes. Nat Immunol, 2002. 3(7): p. 643-51.
229. Lewis, M.D., S.A. Miller, M.M. Miazgowicz, et al., T-bet's ability to regulate individual target genes requires the conserved T-box domain to recruit histone methyltransferase activity and a separate family member-specific transactivation domain. Mol Cell Biol, 2007. 27(24): p. 8510-21.
230. Sekimata, M., M. Perez-Melgosa, S.A. Miller, et al., CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. Immunity, 2009. 31(4): p. 551-64.
231. Ivanov, II, B.S. McKenzie, L. Zhou, et al., The orphan nuclear receptor RORgammat directs the differentiation program of proinflammatory IL-17+ T helper cells. Cell, 2006. 126(6): p. 1121-33.
232. Duhen, R., S. Glatigny, C.A. Arbelaez, et al., Cutting edge: the pathogenicity of IFN-gamma-producing Th17 cells is independent of T-bet. J Immunol, 2013. 190(9): p. 4478-82.
233. Lochner, M., L. Peduto, M. Cherrier, et al., In vivo equilibrium of proinflammatory IL-17+ and regulatory IL-10+ Foxp3+ RORgamma t+ T cells. J Exp Med, 2008. 205(6): p. 1381-93.
234. Tartar, D.M., A.M. VanMorlan, X. Wan, et al., FoxP3+RORgammat+ T helper intermediates display suppressive function against autoimmune diabetes. J Immunol, 2010. 184(7): p. 3377-85.
235. Zhou, X., S.L. Bailey-Bucktrout, L.T. Jeker, et al., Instability of the transcription factor Foxp3 leads to the generation of pathogenic memory T cells in vivo. Nat Immunol, 2009. 10(9): p. 1000-7.
236. Hegazy, A.N., M. Peine, C. Helmstetter, et al., Interferons direct Th2 cell reprogramming to generate a stable GATA-3(+)T-bet(+) cell subset with combined Th2 and Th1 cell functions. Immunity, 2010. 32(1): p. 116-28.
237. Evans, C.M. and R.G. Jenner, Transcription factor interplay in T helper cell differentiation. Brief Funct Genomics, 2013. 12(6): p. 499-511.
238. Stock, P., O. Akbari, G. Berry, et al., Induction of T helper type 1-like regulatory cells that express Foxp3 and protect against airway hyper-reactivity. Nat Immunol, 2004. 5(11): p. 1149-56.
239. Koch, M.A., G. Tucker-Heard, N.R. Perdue, et al., The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. Nat Immunol, 2009. 10(6): p. 595-602.
240. Oldenhove, G., N. Bouladoux, E.A. Wohlfert, et al., Decrease of Foxp3+ Treg cell number and acquisition of effector cell phenotype during lethal infection. Immunity, 2009. 31(5): p. 772-86.
241. Hall, A.O., D.P. Beiting, C. Tato, et al., The cytokines interleukin 27 and interferon-gamma promote distinct Treg cell populations required to limit infection-induced pathology. Immunity, 2012. 37(3): p. 511-23.
242. Rudra, D., P. deRoos, A. Chaudhry, et al., Transcription factor Foxp3 and its protein partners form a complex regulatory network. Nat Immunol, 2012. 13(10): p. 1010-9.
243. Cousins, D.J., T.H. Lee, and D.Z. Staynov, Cytokine coexpression during human Th1/Th2 cell differentiation: direct evidence for coordinated expression of Th2 cytokines. J Immunol, 2002. 169(5): p. 2498-506.
244. Messi, M., I. Giacchetto, K. Nagata, et al., Memory and flexibility of cytokine gene expression as separable properties of human T(H)1 and T(H)2 lymphocytes. Nat Immunol, 2003. 4(1): p. 78-86.
245. Jenner, R.G., M.J. Townsend, I. Jackson, et al., The transcription factors T-bet and GATA-3 control alternative pathways of T-cell differentiation through a shared set of target genes. Proc Natl Acad Sci U S A, 2009. 106(42): p. 17876-81.
246. Kanhere, A., A. Hertweck, U. Bhatia, et al., T-bet and GATA3 orchestrate Th1 and Th2 differentiation through lineage-specific targeting of distal regulatory elements. Nat Commun, 2012. 3: p. 1268.
247. Acosta-Rodriguez, E.V., L. Rivino, J. Geginat, et al., Surface phenotype and

antigenic specificity of human interleukin 17-producing T helper memory cells. Nat Immunol, 2007. 8(6): p. 639-46.
248. Zielinski, C.E., F. Mele, D. Aschenbrenner, et al., Pathogen-induced human TH17 cells produce IFN-gamma or IL-10 and are regulated by IL-1beta. Nature, 2012. 484(7395): p. 514-8.
249. Ghoreschi, K., A. Laurence, X.P. Yang, et al., Generation of pathogenic T(H)17 cells in the absence of TGF-beta signalling. Nature, 2010. 467(7318): p. 967-71.
250. Nistala, K., S. Adams, H. Cambrook, et al., Th17 plasticity in human autoimmune arthritis is driven by the inflammatory environment. Proc Natl Acad Sci U S A, 2010. 107(33): p. 14751-6.
251. Wang, Y.H., K.S. Voo, B. Liu, et al., A novel subset of CD4(+) T(H)2 memory/effector cells that produce inflammatory IL-17 cytokine and promote the exacerbation of chronic allergic asthma. J Exp Med, 2010. 207(11): p. 2479-91.
252. Zhu, J., D. Jankovic, A.J. Oler, et al., The Transcription Factor T-bet Is Induced by Multiple Pathways and Prevents an Endogenous Th2 Cell Program during Th1 Cell Responses. Immunity, 2012. 37(4): p. 660-73.
253. Usui, T., J.C. Preiss, Y. Kanno, et al., T-bet regulates Th1 responses through essential effects on GATA-3 function rather than on IFNG gene acetylation and transcription. J Exp Med, 2006. 203(3): p. 755-66.
254. Garg, V., I.S. Kathiriya, R. Barnes, et al., GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature, 2003. 424(6947): p. 443-7.
255. Oestreich, K.J., A.C. Huang, and A.S. Weinmann, The lineage-defining factors T-bet and Bcl-6 collaborate to regulate Th1 gene expression patterns. J Exp Med, 2011. 208(5): p. 1001-13.
256. Oestreich, K.J., S.E. Mohn, and A.S. Weinmann, Molecular mechanisms that control the expression and activity of Bcl-6 in TH1 cells to regulate flexibility with a TFH-like gene profile. Nat Immunol, 2012. 13(4): p. 405-11.
257. Macian, F., NFAT proteins: key regulators of T-cell development and function. Nat Rev Immunol, 2005. 5(6): p. 472-84.
258. Samstein, R.M., A. Arvey, S.Z. Josefowicz, et al., Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. Cell, 2012. 151(1): p. 153-66.
259. Bettelli, E., M. Dastrange, and M. Oukka, Foxp3 interacts with nuclear factor of activated T cells and NF-kappa B to repress cytokine gene expression and effector functions of T helper cells. Proc Natl Acad Sci U S A, 2005. 102(14): p. 5138-43.
260. Wu, Y., M. Borde, V. Heissmeyer, et al., FOXP3 controls regulatory T cell function through cooperation with NFAT. Cell, 2006. 126(2): p. 375-87.
261. Vahedi, G., H. Takahashi, S. Nakayamada, et al., STATs shape the active enhancer landscape of T cell populations. Cell, 2012. 151(5): p. 981-93.
262. White, S.J., G.H. Underhill, M.H. Kaplan, et al., Cutting edge: differential requirements for Stat4 in expression of glycosyltransferases responsible for selectin ligand formation in Th1 cells. J Immunol, 2001. 167(2): p. 628-31.
263. Hoey, T., S. Zhang, N. Schmidt, et al., Distinct requirements for the naturally occurring splice forms Stat4alpha and Stat4beta in IL-12 responses. EMBO J, 2003. 22(16): p. 4237-48.
264. Thieu, V.T., Q. Yu, H.C. Chang, et al., Signal transducer and activator of transcription 4 is required for the transcription factor T-bet to promote T helper 1 cell-fate determination. Immunity, 2008. 29(5): p. 679-90.
265. Wei, L., G. Vahedi, H.W. Sun, et al., Discrete roles of STAT4 and STAT6 transcription factors in tuning epigenetic modifications and transcription during T helper cell differentiation. Immunity, 2010. 32(6): p. 840-51.
266. Robertson, G., M. Hirst, M. Bainbridge, et al., Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods, 2007. 4(8): p. 651-7.
267. Chen, Z., R. Lund, T. Aittokallio, et al., Identification of novel IL-4/Stat6-regulated genes in T lymphocytes. J Immunol, 2003. 171(7): p. 3627-35.
268. Elo, L.L., H. Jarvenpaa, S. Tuomela, et al., Genome-wide profiling of interleukin-4 and STAT6 transcription factor regulation of human Th2 cell programming. Immunity, 2010. 32(6): p. 852-62.
269. Takeda, K., T. Kaisho, N. Yoshida, et al., Stat3 activation is responsible for IL-6-dependent T cell proliferation through preventing apoptosis: generation and

characterization of T cell-specific Stat3-deficient mice. J Immunol, 1998. 161(9): p. 4652-60.
270. Yang, X.O., A.D. Panopoulos, R. Nurieva, et al., STAT3 regulates cytokine-mediated generation of inflammatory helper T cells. J Biol Chem, 2007. 282(13): p. 9358-63.
271. Burchill, M.A., J. Yang, C. Vogtenhuber, et al., IL-2 receptor beta-dependent STAT5 activation is required for the development of Foxp3+ regulatory T cells. J Immunol, 2007. 178(1): p. 280-90.
272. Chen, C., E.A. Rowell, R.M. Thomas, et al., Transcriptional regulation by Foxp3 is associated with direct promoter occupancy and modulation of histone acetylation. J Biol Chem, 2006. 281(48): p. 36828-34.
273. Zheng, Y., S.Z. Josefowicz, A. Kas, et al., Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells. Nature, 2007. 445(7130): p. 936-40.
274. Grenningloh, R., B.Y. Kang, and I.C. Ho, Ets-1, a functional cofactor of T-bet, is essential for Th1 inflammatory responses. J Exp Med, 2005. 201(4): p. 615-26.
275. Djuretic, I.M., D. Levanon, V. Negreanu, et al., Transcription factors T-bet and Runx3 cooperate to activate Ifng and silence Il4 in T helper type 1 cells. Nat Immunol, 2007. 8(2): p. 145-53.
276. Kohu, K., H. Ohmori, W.F. Wong, et al., The Runx3 transcription factor augments Th1 and down-modulates Th2 phenotypes by interacting with and attenuating GATA3. J Immunol, 2009. 183(12): p. 7817-24.
277. Yagi, R., I.S. Junttila, G. Wei, et al., The transcription factor GATA3 actively represses RUNX3 protein-regulated production of interferon-gamma. Immunity, 2010. 32(4): p. 507-17.
278. Wang, Y., M.A. Su, and Y.Y. Wan, An essential role of the transcription factor GATA-3 for the function of regulatory T cells. Immunity, 2011. 35(3): p. 337-48.
279. Banerji, J., S. Rusconi, and W. Schaffner, Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell, 1981. 27(2 Pt 1): p. 299-308.
280. Moreau, P., R. Hen, B. Wasylyk, et al., The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res, 1981. 9(22): p. 6047-68.
281. Smith, E. and A. Shilatifard, Enhancer biology and enhanceropathies. Nat Struct Mol Biol, 2014. 21(3): p. 210-9.
282. Banerji, J., L. Olson, and W. Schaffner, A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell, 1983. 33(3): p. 729-40.
283. Gillies, S.D., S.L. Morrison, V.T. Oi, et al., A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell, 1983. 33(3): p. 717-28.
284. Mercola, M., X.F. Wang, J. Olsen, et al., Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. Science, 1983. 221(4611): p. 663-5.
285. Klingler, M., J. Soong, B. Butler, et al., Disperse versus compact elements for the regulation of runt stripes in Drosophila. Dev Biol, 1996. 177(1): p. 73-84.
286. Crawford, G.E., I.E. Holt, J. Whittle, et al., Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res, 2006. 16(1): p. 123-31.
287. Stergachis, A.B., S. Neph, A. Reynolds, et al., Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell, 2013. 154(4): p. 888-903.
288. An integrated encyclopedia of DNA elements in the human genome. Nature, 2012. 489(7414): p. 57-74.
289. Heintzman, N.D., G.C. Hon, R.D. Hawkins, et al., Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature, 2009. 459(7243): p. 108-12.
290. Creyghton, M.P., A.W. Cheng, G.G. Welstead, et al., Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A, 2010. 107(50): p. 21931-6.
291. Whyte, W.A., S. Bilodeau, D.A. Orlando, et al., Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature, 2012. 482(7384): p. 221-5.
292. Ogryzko, V.V., R.L. Schiltz, V. Russanova, et al., The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell, 1996. 87(5): p. 953-9.

293. Heintzman, N.D., R.K. Stuart, G. Hon, et al., Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet, 2007. 39(3): p. 311-8.
294. Wang, Z., C. Zang, K. Cui, et al., Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell, 2009. 138(5): p. 1019-31.
295. Visel, A., M.J. Blow, Z. Li, et al., ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature, 2009. 457(7231): p. 854-8.
296. Zhang, W., C. Prakash, C. Sum, et al., Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells. J Biol Chem, 2012. 287(51): p. 43137-55.
297. Dey, A., F. Chitsaz, A. Abbasi, et al., The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. Proc Natl Acad Sci U S A, 2003. 100(15): p. 8758-63.
298. Liu, W., Q. Ma, K. Wong, et al., Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. Cell, 2013. 155(7): p. 1581-95.
299. Panne, D., The enhanceosome. Curr Opin Struct Biol, 2008. 18(2): p. 236-42.
300. Petrij, F., R.H. Giles, H.G. Dauwerse, et al., Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. Nature, 1995. 376(6538): p. 348-51.
301. Roelfsema, J.H., S.J. White, Y. Ariyurek, et al., Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. Am J Hum Genet, 2005. 76(4): p. 572-80.
302. Liu, J. and I.D. Krantz, Cohesin and human disease. Annu Rev Genomics Hum Genet, 2008. 9: p. 303-20.
303. Dorsett, D. and I.D. Krantz, On the molecular etiology of Cornelia de Lange syndrome. Ann N Y Acad Sci, 2009. 1151: p. 22-37.
304. Taub, R., I. Kirsch, C. Morton, et al., Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. Proc Natl Acad Sci U S A, 1982. 79(24): p. 7837-41.
305. Kioussis, D., E. Vanin, T. deLange, et al., Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. Nature, 1983. 306(5944): p. 662-6.
306. Spana, C., D.A. Harrison, and V.G. Corces, The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. Genes Dev, 1988. 2(11): p. 1414-23.
307. Bell, A.C., A.G. West, and G. Felsenfeld, The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell, 1999. 98(3): p. 387-96.
308. Bulger, M. and M. Groudine, Functional and mechanistic diversity of distal transcription enhancers. Cell, 2011. 144(3): p. 327-39.
309. Splinter, E., H. Heath, J. Kooren, et al., CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev, 2006. 20(17): p. 2349-54.
310. Parelho, V., S. Hadjur, M. Spivakov, et al., Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell, 2008. 132(3): p. 422-33.
311. Hirano, T., At the heart of the chromosome: SMC proteins in action. Nat Rev Mol Cell Biol, 2006. 7(5): p. 311-22.
312. Hagstrom, K.A. and B.J. Meyer, Condensin and cohesin: more than chromosome compactor and glue. Nat Rev Genet, 2003. 4(7): p. 520-34.
313. Kagey, M.H., J.J. Newman, S. Bilodeau, et al., Mediator and cohesin connect gene expression and chromatin architecture. Nature, 2010. 467(7314): p. 430-5.
314. Borggrefe, T. and X. Yue, Interactions between subunits of the Mediator complex with gene-specific transcription factors. Semin Cell Dev Biol, 2011. 22(7): p. 759-68.
315. Hwang, Y.C., Q. Zheng, B.D. Gregory, et al., High-throughput identification of long-range regulatory elements and their target promoters in the human genome. Nucleic Acids Res, 2013. 41(9): p. 4835-46.
316. Sanyal, A., B.R. Lajoie, G. Jain, et al., The long-range interaction landscape of gene promoters. Nature, 2012. 489(7414): p. 109-13.
317. He, B., C. Chen, L. Teng, et al., Global view of enhancer-promoter interactome in human cells. Proc Natl Acad Sci U S A, 2014. 111(21): p. E2191-9.
318. Li, Q., K.R. Peterson, X. Fang, et al., Locus control regions. Blood, 2002. 100(9): p. 3077-86.

319. Koch, F., R. Fenouil, M. Gut, et al., Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat Struct Mol Biol, 2011. 18(8): p. 956-63.
320. Loven, J., H.A. Hoke, C.Y. Lin, et al., Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell, 2013. 153(2): p. 320-34.
321. Whyte, W.A., D.A. Orlando, D. Hnisz, et al., Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell, 2013. 153(2): p. 307-19.
322. Moorman, C., L.V. Sun, J. Wang, et al., Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. Proc Natl Acad Sci U S A, 2006. 103(32): p. 12027-32.
323. MacArthur, S., X.Y. Li, J. Li, et al., Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome Biol, 2009. 10(7): p. R80.
324. Roy, S., J. Ernst, P.V. Kharchenko, et al., Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science, 2010. 330(6012): p. 1787-97.
325. Zinzen, R.P., C. Girardot, J. Gagneur, et al., Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature, 2009. 462(7269): p. 65-70.
326. Hnisz, D., B.J. Abraham, T.I. Lee, et al., Super-enhancers in the control of cell identity and disease. Cell, 2013. 155(4): p. 934-47.
327. Ciofani, M., A. Madar, C. Galan, et al., A validated regulatory network for Th17 cell specification. Cell, 2012. 151(2): p. 289-303.
328. Schmidl, C., L. Hansmann, T. Lassmann, et al., The enhancer and promoter landscape of human regulatory and conventional T-cell subpopulations. Blood, 2014. 123(17): p. e68-78.
329. Ouyang, W., O. Beckett, R.A. Flavell, et al., An essential role of the Forkhead-box transcription factor Foxo1 in control of T cell homeostasis and tolerance. Immunity, 2009. 30(3): p. 358-71.
330. Ouyang, W., O. Beckett, Q. Ma, et al., Foxo proteins cooperatively control the differentiation of Foxp3+ regulatory T cells. Nat Immunol, 2010. 11(7): p. 618-27.
331. Agarwal, S. and A. Rao, Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation. Immunity, 1998. 9(6): p. 765-75.
332. Aune, T.M., L.A. Penix, M.R. Rincon, et al., Differential transcription directed by discrete gamma interferon promoter elements in naive and memory (effector) CD4 T cells and CD8 T cells. Mol Cell Biol, 1997. 17(1): p. 199-208.
333. Zhu, H., J. Yang, T.L. Murphy, et al., Unexpected characteristics of the IFN-gamma reporters in nontransformed T cells. J Immunol, 2001. 167(2): p. 855-65.
334. Soutto, M., W. Zhou, and T.M. Aune, Cutting edge: distal regulatory elements are required to achieve selective expression of IFN-gamma in Th1/Tc1 effector cells. J Immunol, 2002. 169(12): p. 6664-7.
335. Hatton, R.D., L.E. Harrington, R.J. Luther, et al., A distal conserved sequence element controls Ifng gene expression by T cells and NK cells. Immunity, 2006. 25(5): p. 717-29.
336. Collins, P.L., S. Chang, M. Henderson, et al., Distal regions of the human IFNG locus direct cell type-specific expression. J Immunol, 2010. 185(3): p. 1492-501.
337. Shnyreva, M., W.M. Weaver, M. Blanchette, et al., Evolutionarily conserved sequence elements that positively regulate IFN-gamma expression in T cells. Proc Natl Acad Sci U S A, 2004. 101(34): p. 12622-7.
338. Lee, D.U., O. Avni, L. Chen, et al., A distal enhancer in the interferon-gamma (IFN-gamma) locus revealed by genome sequence comparison. J Biol Chem, 2004. 279(6): p. 4802-10.
339. Schoenborn, J.R., M.O. Dorschner, M. Sekimata, et al., Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma. Nat Immunol, 2007. 8(7): p. 732-42.
340. Wilson, C.B., E. Rowell, and M. Sekimata, Epigenetic control of T-helper-cell differentiation. Nat Rev Immunol, 2009. 9(2): p. 91-105.
341. Zhang, F. and M. Boothby, T helper type 1-specific Brg1 recruitment and remodeling of nucleosomes positioned at the IFN-gamma promoter are Stat4 dependent. J Exp Med, 2006. 203(6): p. 1493-505.
342. Yang, Y., J.C. Ochando, J.S. Bromberg, et al., Identification of a distant T-bet

enhancer responsive to IL-12/Stat4 and IFNgamma/Stat1 signals. Blood, 2007. 110(7): p. 2494-500.
343. Shi, M., T.H. Lin, K.C. Appell, et al., Janus-kinase-3-dependent signals induce chromatin remodeling at the Ifng locus during T helper 1 cell differentiation. Immunity, 2008. 28(6): p. 763-73.
344. Barski, A., S. Cuddapah, K. Cui, et al., High-resolution profiling of histone methylations in the human genome. Cell, 2007. 129(4): p. 823-37.
345. Wang, Z., C. Zang, J.A. Rosenfeld, et al., Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet, 2008. 40(7): p. 897-903.
346. Boyle, A.P., S. Davis, H.P. Shulha, et al., High-resolution mapping and characterization of open chromatin across the genome. Cell, 2008. 132(2): p. 311-22.
347. Hadjur, S., L.M. Williams, N.K. Ryan, et al., Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature, 2009. 460(7253): p. 410-3.
348. Mujtaba, S., L. Zeng, and M.M. Zhou, Structure and acetyl-lysine recognition of the bromodomain. Oncogene, 2007. 26(37): p. 5521-7.
349. Jang, M.K., K. Mochizuki, M. Zhou, et al., The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol Cell, 2005. 19(4): p. 523-34.
350. Belkina, A.C. and G.V. Denis, BET domain co-regulators in obesity, inflammation and cancer. Nat Rev Cancer, 2012. 12(7): p. 465-77.
351. Kouzarides, T., Chromatin modifications and their function. Cell, 2007. 128(4): p. 693-705.
352. Bernstein, B.E., E.L. Humphrey, R.L. Erlich, et al., Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A, 2002. 99(13): p. 8695-700.
353. Schneider, R., A.J. Bannister, F.A. Myers, et al., Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol, 2004. 6(1): p. 73-7.
354. Li, J., D. Moazed, and S.P. Gygi, Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. J Biol Chem, 2002. 277(51): p. 49383-8.
355. Joshi, A.A. and K. Struhl, Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Mol Cell, 2005. 20(6): p. 971-8.
356. Robertson, A.G., M. Bilenky, A. Tam, et al., Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. Genome Res, 2008. 18(12): p. 1906-17.
357. Muller, J., C.M. Hart, N.J. Francis, et al., Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell, 2002. 111(2): p. 197-208.
358. Wei, G., L. Wei, J. Zhu, et al., Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity, 2009. 30(1): p. 155-67.
359. Tumes, D.J., A. Onodera, A. Suzuki, et al., The polycomb protein Ezh2 regulates differentiation and plasticity of CD4(+) T helper type 1 and type 2 cells. Immunity, 2013. 39(5): p. 819-32.
360. Rea, S., F. Eisenhaber, D. O'Carroll, et al., Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature, 2000. 406(6796): p. 593-9.
361. Le Thomas, A., A.K. Rogers, A. Webster, et al., Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev, 2013. 27(4): p. 390-9.
362. Bulut-Karslioglu, A., I.A. De La Rosa-Velazquez, F. Ramirez, et al., Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Mol Cell, 2014. 55(2): p. 277-90.
363. Vakoc, C.R., S.A. Mandat, B.A. Olenchock, et al., Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell, 2005. 19(3): p. 381-91.
364. Bannister, A.J., P. Zegerman, J.F. Partridge, et al., Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature, 2001. 410(6824): p. 120-4.
365. Margueron, R. and D. Reinberg, The Polycomb complex PRC2 and its mark in life. Nature, 2011. 469(7330): p. 343-9.
366. Tsukiyama, T., C. Daniel, J. Tamkun, et al., ISWI, a member of the SWI2/SNF2

ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. Cell, 1995. 83(6): p. 1021-6.
367. Badenhorst, P., M. Voas, I. Rebay, et al., Biological functions of the ISWI chromatin remodeling complex NURF. Genes Dev, 2002. 16(24): p. 3186-98.
368. Wysocka, J., T. Swigut, T.A. Milne, et al., WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell, 2005. 121(6): p. 859-72.
369. Azuara, V., P. Perry, S. Sauer, et al., Chromatin signatures of pluripotent cell lines. Nat Cell Biol, 2006. 8(5): p. 532-8.
370. Fischle, W., Y. Wang, S.A. Jacobs, et al., Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev, 2003. 17(15): p. 1870-81.
371. Saint-Andre, V., E. Batsche, C. Rachez, et al., Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alternative exons. Nat Struct Mol Biol, 2011. 18(3): p. 337-44.
372. Squazzo, S.L., H. O'Geen, V.M. Komashko, et al., Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res, 2006. 16(7): p. 890-900.
373. Kornberg, R.D., The molecular basis of eukaryotic transcription. Proc Natl Acad Sci U S A, 2007. 104(32): p. 12955-61.
374. Rougvie, A.E. and J.T. Lis, The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell, 1988. 54(6): p. 795-804.
375. Adelman, K. and J.T. Lis, Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet, 2012. 13(10): p. 720-31.
376. Peng, J., N.F. Marshall, and D.H. Price, Identification of a cyclin subunit required for the function of Drosophila P-TEFb. J Biol Chem, 1998. 273(22): p. 13855-60.
377. Yang, Z., Q. Zhu, K. Luo, et al., The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature, 2001. 414(6861): p. 317-22.
378. Nguyen, V.T., T. Kiss, A.A. Michels, et al., 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature, 2001. 414(6861): p. 322-5.
379. Luo, Z., C. Lin, and A. Shilatifard, The super elongation complex (SEC) family in transcriptional control. Nat Rev Mol Cell Biol, 2012. 13(9): p. 543-7.
380. Chao, S.H. and D.H. Price, Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. J Biol Chem, 2001. 276(34): p. 31793-9.
381. Filippakopoulos, P., J. Qi, S. Picaud, et al., Selective inhibition of BET bromodomains. Nature, 2010. 468(7327): p. 1067-73.
382. Chapuy, B., M.R. McKeown, C.Y. Lin, et al., Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell, 2013. 24(6): p. 777-90.
383. Brown, J.D., C.Y. Lin, Q. Duan, et al., NF-kappaB Directs Dynamic Super Enhancer Formation in Inflammation and Atherogenesis. Mol Cell, 2014. 56(2): p. 219-31.
384. Di Micco, R., B. Fontanals-Cirera, V. Low, et al., Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. Cell Rep, 2014. 9(1): p. 234-47.
385. Kaikkonen, M.U., N.J. Spann, S. Heinz, et al., Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol Cell, 2013. 51(3): p. 310-25.
386. De Santa, F., I. Barozzi, F. Mietton, et al., A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol, 2010. 8(5): p. e1000384.
387. Hah, N., S. Murakami, A. Nagari, et al., Enhancer transcripts mark active estrogen receptor binding sites. Genome Res, 2013. 23(8): p. 1210-23.
388. Rogelj, B. and K.P. Giese, Expression and function of brain specific small RNAs. Rev Neurosci, 2004. 15(3): p. 185-98.
389. Bratkovic, T. and B. Rogelj, Biology and applications of small nucleolar RNAs. Cell Mol Life Sci, 2011. 68(23): p. 3843-51.
390. Bartel, D.P., MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 2004. 116(2): p. 281-97.
391. Siomi, M.C., K. Sato, D. Pezic, et al., PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol, 2011. 12(4): p. 246-58.

392. Lander, E.S., L.M. Linton, B. Birren, et al., Initial sequencing and analysis of the human genome. Nature, 2001. 409(6822): p. 860-921.
393. Wong, G.K., D.A. Passey, and J. Yu, Most of the human genome is transcribed. Genome Res, 2001. 11(12): p. 1975-7.
394. Hangauer, M.J., I.W. Vaughn, and M.T. McManus, Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. PLoS Genet, 2013. 9(6): p. e1003569.
395. Carninci, P., T. Kasukawa, S. Katayama, et al., The transcriptional landscape of the mammalian genome. Science, 2005. 309(5740): p. 1559-63.
396. Struhl, K., Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol, 2007. 14(2): p. 103-5.
397. Graur, D., Y. Zheng, N. Price, et al., On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. Genome Biol Evol, 2013. 5(3): p. 578-90.
398. Zhao, J., B.K. Sun, J.A. Erwin, et al., Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science, 2008. 322(5902): p. 750-6.
399. Esteller, M., Non-coding RNAs in human disease. Nat Rev Genet, 2011. 12(12): p. 861-74.
400. MirVana miRNA Isolation Kit. Ambion Life Technologies, 2011.
401. McSweeney, M. Long noncoding RNA: The dark matter of the genome. Accessed November 2014; Available from: http://mcmanuslab.ucsf.edu/sites/mcmanuslab.ucsf.edu/files/presentations/MM_lncRNA%20presentation.pdf.
402. Andrews, S.J. and J.A. Rothnagel, Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet, 2014. 15(3): p. 193-204.
403. Derrien, T., R. Johnson, G. Bussotti, et al., The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res, 2012. 22(9): p. 1775-89.
404. Ulitsky, I. and D.P. Bartel, lincRNAs: genomics, evolution, and mechanisms. Cell, 2013. 154(1): p. 26-46.
405. Guttman, M., I. Amit, M. Garber, et al., Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature, 2009. 458(7235): p. 223-7.
406. Kutter, C., S. Watt, K. Stefflova, et al., Rapid turnover of long noncoding RNAs and the evolution of gene expression. PLoS Genet, 2012. 8(7): p. e1002841.
407. Kelley, D. and J. Rinn, Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol, 2012. 13(11): p. R107.
408. Ravasi, T., H. Suzuki, K.C. Pang, et al., Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res, 2006. 16(1): p. 11-9.
409. Schorderet, P. and D. Duboule, Structural and functional differences in the long non-coding RNA hotair in mouse and human. PLoS Genet, 2011. 7(5): p. e1002071.
410. Sauvageau, M., L.A. Goff, S. Lodato, et al., Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife, 2013. 2: p. e01749.
411. Kapranov, P., J. Cheng, S. Dike, et al., RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science, 2007. 316(5830): p. 1484-8.
412. Seila, A.C., J.M. Calabrese, S.S. Levine, et al., Divergent transcription from active promoters. Science, 2008. 322(5909): p. 1849-51.
413. Core, L.J., J.J. Waterfall, and J.T. Lis, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science, 2008. 322(5909): p. 1845-8.
414. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature, 2009. 457(7232): p. 1028-32.
415. Taft, R.J., E.A. Glazov, N. Cloonan, et al., Tiny RNAs associated with transcription start sites in animals. Nat Genet, 2009. 41(5): p. 572-8.
416. Kanhere, A., K. Viiri, C.C. Araujo, et al., Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol Cell, 2010. 38(5): p. 675-88.
417. Preker, P., J. Nielsen, S. Kammler, et al., RNA exosome depletion reveals transcription upstream of active human promoters. Science, 2008. 322(5909): p. 1851-4.

418. Ntini, E., A.I. Jarvelin, J. Bornholdt, et al., Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nat Struct Mol Biol, 2013. 20(8): p. 923-8.
419. Jacquier, A., The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat Rev Genet, 2009. 10(12): p. 833-44.
420. Kim, T.K., M. Hemberg, J.M. Gray, et al., Widespread transcription at neuronal activity-regulated enhancers. Nature, 2010. 465(7295): p. 182-7.
421. Ilott, N.E., J.A. Heward, B. Roux, et al., Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. Nat Commun, 2014. 5: p. 3979.
422. Melgar, M.F., F.S. Collins, and P. Sethupathy, Discovery of active enhancers through bidirectional expression of short transcripts. Genome Biol, 2011. 12(11): p. R113.
423. Andersson, R., C. Gebhard, I. Miguel-Escalada, et al., An atlas of active enhancers across human cell types and tissues. Nature, 2014. 507(7493): p. 455-61.
424. Ashe, H.L., J. Monks, M. Wijgerde, et al., Intergenic transcription and transinduction of the human beta-globin locus. Genes Dev, 1997. 11(19): p. 2494-509.
425. Li, W., D. Notani, Q. Ma, et al., Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature, 2013. 498(7455): p. 516-20.
426. Lai, F., U.A. Orom, M. Cesaroni, et al., Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature, 2013.
427. Schaukowitch, K., J.Y. Joo, X. Liu, et al., Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell, 2014. 56(1): p. 29-42.
428. Mousavi, K., H. Zare, S. Dell'orso, et al., eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell, 2013. 51(5): p. 606-17.
429. Lam, M.T., H. Cho, H.P. Lesch, et al., Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature, 2013. 498(7455): p. 511-5.
430. Rinn, J.L. and H.Y. Chang, Genome regulation by long noncoding RNAs. Annu Rev Biochem, 2012. 81: p. 145-66.
431. Wang, X., S. Arai, X. Song, et al., Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature, 2008. 454(7200): p. 126-30.
432. Guil, S. and M. Esteller, Cis-acting noncoding RNAs: friends and foes. Nat Struct Mol Biol, 2012. 19(11): p. 1068-75.
433. Hacisuleyman, E., L.A. Goff, C. Trapnell, et al., Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat Struct Mol Biol, 2014. 21(2): p. 198-206.
434. Preker, P., K. Almvig, M.S. Christensen, et al., PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. Nucleic Acids Res, 2011. 39(16): p. 7179-93.
435. Novikova, I.V., S.P. Hennelly, C.S. Tung, et al., Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs. J Mol Biol, 2013.
436. Gomez, J.A., O.L. Wapinski, Y.W. Yang, et al., The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. Cell, 2013. 152(4): p. 743-54.
437. Turner, M., A. Galloway, and E. Vigorito, Noncoding RNA and its associated proteins as regulatory elements of the immune system. Nat Immunol, 2014. 15(6): p. 484-91.
438. Imamura, T., S. Yamamoto, J. Ohgane, et al., Non-coding RNA directed DNA demethylation of Sphk1 CpG island. Biochem Biophys Res Commun, 2004. 322(2): p. 593-600.
439. Sabin, L.R., M.J. Delas, and G.J. Hannon, Dogma derailed: the many influences of RNA on the genome. Mol Cell, 2013. 49(5): p. 783-94.
440. Yao, H., K. Brick, Y. Evrard, et al., Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. Genes Dev, 2010. 24(22): p. 2543-55.
441. Khalil, A.M., M. Guttman, M. Huarte, et al., Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A, 2009. 106(28): p. 11667-72.
442. Guttman, M., J. Donaghey, B.W. Carey, et al., lincRNAs act in the circuitry

controlling pluripotency and differentiation. Nature, 2011. 477(7364): p. 295-300.
443. Brockdorff, N., Noncoding RNA and Polycomb recruitment. RNA, 2013. 19(4): p. 429-42.
444. Hafner, M., M. Landthaler, L. Burger, et al., Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell, 2010. 141(1): p. 129-41.
445. Kaneko, S., R. Bonasio, R. Saldana-Meyer, et al., Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. Mol Cell, 2014. 53(2): p. 290-300.
446. Wutz, A. and R. Jaenisch, A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell, 2000. 5(4): p. 695-705.
447. Sun, B.K., A.M. Deaton, and J.T. Lee, A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol Cell, 2006. 21(5): p. 617-28.
448. Beltran, M., I. Puig, C. Pena, et al., A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev, 2008. 22(6): p. 756-69.
449. Orom, U.A., T. Derrien, M. Beringer, et al., Long noncoding RNAs with enhancer-like function in human cells. Cell, 2010. 143(1): p. 46-58.
450. Taatjes, D.J., The human Mediator complex: a versatile, genome-wide regulator of transcription. Trends Biochem Sci, 2010. 35(6): p. 315-22.
451. Wang, K.C., Y.W. Yang, B. Liu, et al., A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature, 2011. 472(7341): p. 120-4.
452. Kmita, M. and D. Duboule, Organizing axes in time and space; 25 years of colinear tinkering. Science, 2003. 301(5631): p. 331-3.
453. Beuchle, D., G. Struhl, and J. Muller, Polycomb group proteins and heritable silencing of Drosophila Hox genes. Development, 2001. 128(6): p. 993-1004.
454. Rinn, J.L., M. Kertesz, J.K. Wang, et al., Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell, 2007. 129(7): p. 1311-23.
455. Tsai, M.C., O. Manor, Y. Wan, et al., Long noncoding RNA as modular scaffold of histone modification complexes. Science, 2010. 329(5992): p. 689-93.
456. Gupta, R.A., N. Shah, K.C. Wang, et al., Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature, 2010. 464(7291): p. 1071-6.
457. Tripathi, V., J.D. Ellis, Z. Shen, et al., The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell, 2010. 39(6): p. 925-38.
458. Eissmann, M., T. Gutschner, M. Hammerle, et al., Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. RNA Biol, 2012. 9(8): p. 1076-87.
459. Kino, T., D.E. Hurt, T. Ichijo, et al., Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal, 2010. 3(107): p. ra8.
460. Hung, T., Y. Wang, M.F. Lin, et al., Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet, 2011. 43(7): p. 621-9.
461. Hu, G., Q. Tang, S. Sharma, et al., Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. Nat Immunol, 2013.
462. Willingham, A.T., A.P. Orth, S. Batalov, et al., A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science, 2005. 309(5740): p. 1570-3.
463. Vigneau, S., P.S. Rohrlich, M. Brahic, et al., Tmevpg1, a candidate gene for the control of Theiler's virus persistence, could be implicated in the regulation of gamma interferon. J Virol, 2003. 77(10): p. 5632-8.
464. Collier, S.P., P.L. Collins, C.L. Williams, et al., Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of Ifng by Th1 cells. J Immunol, 2012. 189(5): p. 2084-8.
465. Collier, S.P., M.A. Henderson, J.T. Tossberg, et al., Regulation of the Th1 Genomic Locus from Ifng through Tmevpg1 by T-bet. J Immunol, 2014. 193(8): p. 3959-65.
466. Rothstein, D.M., A. Yamada, S.F. Schlossman, et al., Cyclic regulation of CD45

isoform expression in a long term human CD4+CD45RA+ T cell line. J Immunol, 1991. 146(4): p. 1175-83.
467. algorithm, D.o.p.s.d., http://dharmacon.gelifesciences.com/design-center/.
468. Mantei, A., S. Rutz, M. Janke, et al., siRNA stabilization prolongs gene knockdown in primary T lymphocytes. Eur J Immunol, 2008. 38(9): p. 2616-25.
469. Amaxa Human T cell Nucleafector kit, VPA-1002, Lonza.
470. Exiqon LNA design algorithm. http://www.exiqon.com/ls/Pages/GDTSequenceInput.aspx.
471. Technologies, A., Agilent RNA 6000 Pico Kit Guide. 2008.
472. Two-Color Microarray-Based Gene Expression analysis; Low Input Quick Amp Labeling. Agilent Technologies, 2010(Version 6.5).
473. Roche, RealTime ready. Universal ProbeLibrary. Redefining and revolutionizing real-time qPCR assays. 2009.
474. http://www.roche-applied-science.com/sis/rtpcr/upl/index.jsp?id=UP030000.
475. http://primer3.ut.ee/, P.
476. de Hoon, M.J., S. Imoto, J. Nolan, et al., Open source clustering software. Bioinformatics, 2004. 20(9): p. 1453-4.
477. Saldanha, A.J., Java Treeview--extensible visualization of microarray data. Bioinformatics, 2004. 20(17): p. 3246-8.
478. Laing, E. and C.P. Smith, RankProdIt: A web-interactive Rank Products analysis tool. BMC Res Notes, 2010. 3: p. 221.
479. Breitling, R., P. Armengaud, A. Amtmann, et al., Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett, 2004. 573(1-3): p. 83-92.
480. Hong, F., R. Breitling, C.W. McEntee, et al., RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics, 2006. 22(22): p. 2825-7.
481. http://strep-microarray.sbs.surrey.ac.uk/RankProducts/.
482. http://genome.ucsc.edu/.
483. Benson, D.A., M. Cavanaugh, K. Clark, et al., GenBank. Nucleic Acids Res, 2013. 41(Database issue): p. D36-42.
484. Pruitt, K.D., J. Harrow, R.A. Harte, et al., The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome Res, 2009. 19(7): p. 1316-23.
485. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res, 2014. 42(Database issue): p. D191-8.
486. Griffiths-Jones, S., A. Bateman, M. Marshall, et al., Rfam: an RNA family database. Nucleic Acids Res, 2003. 31(1): p. 439-41.
487. http://genome.ucsc.edu/cgi-bin/hgGene?hgg_do_kgMethod=1&hgsid=320247085.
488. Flicek, P., M.R. Amode, D. Barrell, et al., Ensembl 2011. Nucleic Acids Res, 2011. 39(Database issue): p. D800-6.
489. Harrow, J., F. Denoeud, A. Frankish, et al., GENCODE: producing a reference annotation for ENCODE. Genome Biol, 2006. 7 Suppl 1: p. S4 1-9.
490. Pruitt, K.D., T. Tatusova, and D.R. Maglott, NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res, 2005. 33(Database issue): p. D501-4.
491. http://www.ncbi.nlm.nih.gov/RefSeq/key.html#status.
492. Kent, W.J., BLAT--the BLAST-like alignment tool. Genome Res, 2002. 12(4): p. 656-64.
493. http://vega.sanger.ac.uk/info/about/gene_and_transcript_types.html.
494. Ensembl lincRNA annotation guidlines. http://www.ensembl.org/info/genome/genebuild/ncrna.html.
495. Hubbard, T., D. Barker, E. Birney, et al., The Ensembl genome database project. Nucleic Acids Res, 2002. 30(1): p. 38-41.
496. Cabili, M.N., C. Trapnell, L. Goff, et al., Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev, 2011. 25(18): p. 1915-27.
497. Kozomara, A. and S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res, 2011. 39(Database issue): p. D152-7.
498. Lestrade, L. and M.J. Weber, snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic Acids Res, 2006. 34(Database

issue): p. D158-62.
499. Huang da, W., B.T. Sherman, and R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc, 2009. 4(1): p. 44-57.
500. Huang da, W., B.T. Sherman, and R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res, 2009. 37(1): p. 1-13.
501. Becker, K.G., K.C. Barnes, T.J. Bright, et al., The genetic association database. Nat Genet, 2004. 36(5): p. 431-2.
502. Hamosh, A., A.F. Scott, J.S. Amberger, et al., Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res, 2005. 33(Database issue): p. D514-7.
503. Landrum, M.J., J.M. Lee, G.R. Riley, et al., ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res, 2014. 42(Database issue): p. D980-5.
504. Miller, D.T., M.P. Adam, S. Aradhya, et al., Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet, 2010. 86(5): p. 749-64.
505. Forbes, S.A., G. Bhamra, S. Bamford, et al., The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet, 2008. Chapter 10: p. Unit 10 11.
506. Laulederkind, S.J., G.T. Hayman, S.J. Wang, et al., The Rat Genome Database 2013--data, tools and users. Brief Bioinform, 2013. 14(4): p. 520-6.
507. Stenson, P.D., E.V. Ball, K. Howells, et al., The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. Hum Genomics, 2009. 4(2): p. 69-72.
508. Roberta A Pagon, M.P.A., Holly H Ardinger, Thomas D Bird, Cynthia R Dolan, Chin-To Fong, Richard JH Smith, and Karen Stephens, GeneReviews1993-2014, Seattle (WA): University of Washington, Seattle.
509. Hindorff, L.A., P. Sethupathy, H.A. Junkins, et al., Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A, 2009. 106(23): p. 9362-7.
510. Vaquerizas, J.M., S.K. Kummerfeld, S.A. Teichmann, et al., A census of human transcription factors: function, expression and evolution. Nat Rev Genet, 2009. 10(4): p. 252-63.
511. Li, H., B. Handsaker, A. Wysoker, et al., The Sequence Alignment/Map format and SAMtools. Bioinformatics, 2009. 25(16): p. 2078-9.
512. Barnett, D.W., E.K. Garrison, A.R. Quinlan, et al., BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics, 2011. 27(12): p. 1691-2.
513. Trapnell, C., D.G. Hendrickson, M. Sauvageau, et al., Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol, 2013. 31(1): p. 46-53.
514. Kolmogorov-Smirnov Test: http://www.physics.csbsju.edu/stats/KS-test.html.
515. Zhang, Y., T. Liu, C.A. Meyer, et al., Model-based analysis of ChIP-Seq (MACS). Genome Biol, 2008. 9(9): p. R137.
516. FASTQC, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
517. Birzele, F., T. Fauti, H. Stahl, et al., Next-generation insights into regulatory T cells: expression profiling and FoxP3 occupancy in Human. Nucleic Acids Res, 2011. 39(18): p. 7946-60.
518. Read, S., V. Malmstrom, and F. Powrie, Cytotoxic T lymphocyte-associated antigen 4 plays an essential role in the function of CD25(+)CD4(+) regulatory cells that control intestinal inflammation. J Exp Med, 2000. 192(2): p. 295-302.
519. Demeure, C.E., D.G. Byun, L.P. Yang, et al., CD31 (PECAM-1) is a differentiation antigen lost during human CD4 T-cell maturation into Th1 or Th2 effector cells. Immunology, 1996. 88(1): p. 110-5.
520. Cobb, B.S., A. Hertweck, J. Smith, et al., A role for Dicer in immune regulation. J Exp Med, 2006. 203(11): p. 2519-27.
521. Lu, L.F., M.P. Boldin, A. Chaudhry, et al., Function of miR-146a in controlling Treg cell-mediated regulation of Th1 responses. Cell, 2010. 142(6): p. 914-29.
522. Pandey, R.R., T. Mondal, F. Mohammad, et al., Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell, 2008. 32(2): p. 232-46.
523. Wutz, A., T.P. Rasmussen, and R. Jaenisch, Chromosomal silencing and

localization are mediated by different domains of Xist RNA. Nat Genet, 2002. 30(2): p. 167-74.
524. Graham, L.D., S.K. Pedersen, G.S. Brown, et al., Colorectal Neoplasia Differentially Expressed (CRNDE), a Novel Gene with Elevated Expression in Colorectal Adenomas and Adenocarcinomas. Genes Cancer, 2011. 2(8): p. 829-40.
525. Wolfs, M.G., S.S. Rensen, E.J. Bruin-Van Dijk, et al., Co-expressed immune and metabolic genes in visceral and subcutaneous adipose tissue from severely obese individuals are associated with plasma HDL and glucose levels: a microarray study. BMC Med Genomics, 2010. 3: p. 34.
526. Ellis, B.C., P.L. Molloy, and L.D. Graham, CRNDE: A Long Non-Coding RNA Involved in CanceR, Neurobiology, and DEvelopment. Front Genet, 2012. 3: p. 270.
527. Cheng, C.W., R.L. Chow, M. Lebel, et al., The Iroquois homeobox gene, Irx5, is required for retinal cone bipolar cell development. Dev Biol, 2005. 287(1): p. 48-60.
528. Gaborit, N., R. Sakuma, J.N. Wylie, et al., Cooperative and antagonistic roles for Irx3 and Irx5 in cardiac morphogenesis and postnatal physiology. Development, 2012. 139(21): p. 4007-19.
529. Bian, Z.Y., H. Huang, H. Jiang, et al., LIM and cysteine-rich domains 1 regulates cardiac hypertrophy by targeting calcineurin/nuclear factor of activated T cells signaling. Hypertension, 2010. 55(2): p. 257-63.
530. Liu, M., L. Chen, T.H. Chan, et al., Serum and glucocorticoid kinase 3 at 8q13.1 promotes cell proliferation and survival in hepatocellular carcinoma. Hepatology, 2012. 55(6): p. 1754-65.
531. Kuznetsova, E.B., T.V. Kekeeva, S.S. Larin, et al., [Novel methylation and expression markers associated with breast cancer]. Mol Biol (Mosk), 2007. 41(4): p. 624-33.
532. Sheleg, S.V., J.M. Peloponese, Y.H. Chi, et al., Evidence for cooperative transforming activity of the human pituitary tumor transforming gene and human T-cell leukemia virus type 1 Tax. J Virol, 2007. 81(15): p. 7894-901.
533. Zhang, G., Q. Zhao, S. Yu, et al., Pttg1 inhibits TGFbeta signaling in breast cancer cells to promote their growth. Tumour Biol, 2014.
534. Gomes, A.Q., D.V. Correia, A.R. Grosso, et al., Identification of a panel of ten cell surface protein antigens associated with immunotargeting of leukemias and lymphomas by peripheral blood gammadelta T cells. Haematologica, 2010. 95(8): p. 1397-404.
535. Jawaheer, D., M.F. Seldin, C.I. Amos, et al., A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. Am J Hum Genet, 2001. 68(4): p. 927-36.
536. Rapetti, L., K.M. Chavele, C.M. Evans, et al., B cell resistance to Fas-mediated apoptosis contributes to their ineffective control by regulatory T cells in rheumatoid arthritis. Ann Rheum Dis, 2013.
537. Yokouchi, Y., Y. Nukaga, M. Shibasaki, et al., Significant evidence for linkage of mite-sensitive childhood asthma to chromosome 5q31-q33 near the interleukin 12 B locus by a genome-wide search in Japanese families. Genomics, 2000. 66(2): p. 152-60.
538. Xu, B., N.H. Feng, P.C. Li, et al., A functional polymorphism in Pre-miR-146a gene is associated with prostate cancer risk and mature miR-146a expression in vivo. Prostate, 2010. 70(5): p. 467-72.
539. Jazdzewski, K., S. Liyanarachchi, M. Swierniak, et al., Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. Proc Natl Acad Sci U S A, 2009. 106(5): p. 1502-5.
540. Luo, X., W. Yang, D.Q. Ye, et al., A functional variant in microRNA-146a promoter modulates its expression and confers disease risk for systemic lupus erythematosus. PLoS Genet, 2011. 7(6): p. e1002128.
541. Ishii, T., H. Onda, A. Tanigawa, et al., Isolation and expression profiling of genes upregulated in the peripheral blood cells of systemic lupus erythematosus patients. DNA Res, 2005. 12(6): p. 429-39.
542. Hua, J., K. Kirou, C. Lee, et al., Functional assay of type I interferon in systemic lupus erythematosus plasma and association with anti-RNA binding protein autoantibodies. Arthritis Rheum, 2006. 54(6): p. 1906-16.
543. Nzeusseu Toukap, A., C. Galant, I. Theate, et al., Identification of distinct gene

expression profiles in the synovium of patients with systemic lupus erythematosus. Arthritis Rheum, 2007. 56(5): p. 1579-88.
544. Han, G.M., S.L. Chen, N. Shen, et al., Analysis of gene expression profiles in human systemic lupus erythematosus using oligonucleotide microarray. Genes Immun, 2003. 4(3): p. 177-86.
545. Hashizume, H., H. Hamalainen, Q. Sun, et al., Downregulation of mafB expression in T-helper cells during early differentiation in vitro. Scand J Immunol, 2003. 57(1): p. 28-34.
546. Gemelli, C., T. Zanocco Marani, S. Bicciato, et al., MafB is a downstream target of the IL-10/STAT3 signaling pathway, involved in the regulation of macrophage de-activation. Biochim Biophys Acta, 2014. 1843(5): p. 955-64.
547. Sekiya, T., I. Kashiwagi, R. Yoshida, et al., Nr4a receptors are essential for thymic regulatory T cell development and immune homeostasis. Nat Immunol, 2013. 14(3): p. 230-7.
548. Chang, Q., J. Chen, K.J. Beezhold, et al., JNK1 activation predicts the prognostic outcome of the human hepatocellular carcinoma. Mol Cancer, 2009. 8: p. 64.
549. Lin, M., E. Pedrosa, A. Shah, et al., RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. PLoS One, 2011. 6(9): p. e23356.
550. Ellis, B.C., L.D. Graham, and P.L. Molloy, CRNDE, a long non-coding RNA responsive to insulin/IGF signaling, regulates genes involved in central metabolism. Biochim Biophys Acta, 2014. 1843(2): p. 372-86.
551. Lee, M.S., K. Hanspers, C.S. Barker, et al., Gene expression profiles during human CD4+ T cell differentiation. Int Immunol, 2004. 16(8): p. 1109-24.
552. Davidovich, C., L. Zheng, K.J. Goodrich, et al., Promiscuous RNA binding by Polycomb repressive complex 2. Nat Struct Mol Biol, 2013. 20(11): p. 1250-7.
553. Cerase, A., D. Smeets, Y.A. Tang, et al., Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. Proc Natl Acad Sci U S A, 2014. 111(6): p. 2235-40.
554. Ho, I.C., D. Lo, and L.H. Glimcher, c-maf promotes T helper cell type 2 (Th2) and attenuates Th1 differentiation by both interleukin 4-dependent and -independent mechanisms. J Exp Med, 1998. 188(10): p. 1859-66.
555. Naito, T., H. Tanaka, Y. Naoe, et al., Transcriptional control of T-cell development. Int Immunol, 2011. 23(11): p. 661-8.
556. Tanaka, S., A. Suto, T. Iwamoto, et al., Sox5 and c-Maf cooperatively induce Th17 cell differentiation via RORgammat induction as downstream targets of Stat3. J Exp Med, 2014. 211(9): p. 1857-74.
557. Sieweke, M.H., H. Tekotte, J. Frampton, et al., MafB is an interaction partner and repressor of Ets-1 that inhibits erythroid differentiation. Cell, 1996. 85(1): p. 49-60.
558. Natoli, G. and J.C. Andrau, Noncoding transcription at enhancers: general principles and functional models. Annu Rev Genet, 2012. 46: p. 1-19.
559. Orom, U.A. and R. Shiekhattar, Long noncoding RNAs usher in a new era in the biology of enhancers. Cell, 2013. 154(6): p. 1190-3.
560. Lam, M.T., W. Li, M.G. Rosenfeld, et al., Enhancer RNAs and regulated transcriptional programs. Trends Biochem Sci, 2014. 39(4): p. 170-82.
561. Kreutzman, A., M. Ilander, K. Porkka, et al., Dasatinib promotes Th1-type responses in granzyme B expressing T-cells. Oncoimmunology, 2014. 3: p. e28925.
562. Pesu, M., L. Muul, Y. Kanno, et al., Proprotein convertase furin is preferentially expressed in T helper 1 cells and regulates interferon gamma. Blood, 2006. 108(3): p. 983-5.
563. Keskin, H., J. Garriga, D. Georlette, et al., Complex effects of flavopiridol on the expression of primary response genes. Cell Div, 2012. 7: p. 11.
564. Pirngruber, J., A. Shchebet, and S.A. Johnsen, Insights into the function of the human P-TEFb component CDK9 in the regulation of chromatin modifications and co-transcriptional mRNA processing. Cell Cycle, 2009. 8(22): p. 3636-42.
565. Lee, Y., K. Jeon, J.T. Lee, et al., MicroRNA maturation: stepwise processing and subcellular localization. EMBO J, 2002. 21(17): p. 4663-70.
566. Lee, Y., C. Ahn, J. Han, et al., The nuclear RNase III Drosha initiates microRNA processing. Nature, 2003. 425(6956): p. 415-9.
567. Yi, R., Y. Qin, I.G. Macara, et al., Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev, 2003. 17(24): p. 3011-6.
568. Bernstein, E., A.A. Caudy, S.M. Hammond, et al., Role for a bidentate

ribonuclease in the initiation step of RNA interference. Nature, 2001. 409(6818): p. 363-6.
569. Meister, G., M. Landthaler, A. Patkaniowska, et al., Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell, 2004. 15(2): p. 185-97.
570. Sontheimer, E.J., Assembly and function of RNA silencing complexes. Nat Rev Mol Cell Biol, 2005. 6(2): p. 127-38.
571. Liu, J., Control of protein synthesis and mRNA degradation by microRNAs. Curr Opin Cell Biol, 2008. 20(2): p. 214-21.
572. Gibbings, D.J., C. Ciaudo, M. Erhardt, et al., Multivesicular bodies associate with components of miRNA effector complexes and modulate miRNA activity. Nat Cell Biol, 2009. 11(9): p. 1143-9.
573. Robb, G.B., K.M. Brown, J. Khurana, et al., Specific and potent RNAi in the nucleus of human cells. Nat Struct Mol Biol, 2005. 12(2): p. 133-7.
574. Provost, P., D. Dishart, J. Doucet, et al., Ribonuclease activity and RNA binding of recombinant human Dicer. EMBO J, 2002. 21(21): p. 5864-74.
575. Wengel, J., M. Petersen, K.E. Nielsen, et al., LNA (locked nucleic acid) and the diastereoisomeric alpha-L-LNA: conformational tuning and high-affinity recognition of DNA/RNA targets. Nucleosides Nucleotides Nucleic Acids, 2001. 20(4-7): p. 389-96.
576. Kaur, H., A. Arora, J. Wengel, et al., Thermodynamic, counterion, and hydration effects for the incorporation of locked nucleic acid nucleotides into DNA duplexes. Biochemistry, 2006. 45(23): p. 7347-55.
577. Braasch, D.A. and D.R. Corey, Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. Chem Biol, 2001. 8(1): p. 1-7.
578. Zamaratski, E., P.I. Pradeepkumar, and J. Chattopadhyaya, A critical survey of the structure-function of the antisense oligo/RNA heteroduplex as substrate for RNase H. J Biochem Biophys Methods, 2001. 48(3): p. 189-208.
579. Arzumanov, A., D.A. Stetsenko, A.D. Malakhov, et al., A structure-activity study of the inhibition of HIV-1 Tat-dependent trans-activation by mixmer 2'-O-methyl oligoribonucleotides containing locked nucleic acid (LNA), alpha-L-LNA, or 2'-thio-LNA residues. Oligonucleotides, 2003. 13(6): p. 435-53.
580. Stein, C.A., J.B. Hansen, J. Lai, et al., Efficient gene silencing by delivery of locked nucleic acid antisense oligonucleotides, unassisted by transfection reagents. Nucleic Acids Res, 2010. 38(1): p. e3.
581. Zhu, T., Z. Dou, B. Qin, et al., Phosphorylation of microtubule-binding protein Hec1 by mitotic kinase Aurora B specifies spindle checkpoint kinase Mps1 signaling at the kinetochore. J Biol Chem, 2013. 288(50): p. 36149-59.
582. Merkle, C.J., L.M. Karnitz, J.T. Henry-Sanchez, et al., Cloning and characterization of hCTF18, hCTF8, and hDCC1. Human homologs of a Saccharomyces cerevisiae complex involved in sister chromatid cohesion establishment. J Biol Chem, 2003. 278(32): p. 30051-6.
583. Dai, Z., Q. Li, Y. Wang, et al., CD4+CD25+ regulatory T cells suppress allograft rejection mediated by memory CD8+ T cells via a CD30-dependent mechanism. J Clin Invest, 2004. 113(2): p. 310-7.
584. Elpek, K.G., E.S. Yolcu, D.D. Franke, et al., Ex vivo expansion of CD4+CD25+FoxP3+ T regulatory cells based on synergy between IL-2 and 4-1BB signaling. J Immunol, 2007. 179(11): p. 7295-304.
585. Zhang, P., F. Gao, Q. Wang, et al., Agonistic anti-4-1BB antibody promotes the expansion of natural regulatory T cells while maintaining Foxp3 expression. Scand J Immunol, 2007. 66(4): p. 435-40.
586. Lopez, F., F. Belloc, F. Lacombe, et al., Modalities of synthesis of Ki67 antigen during the stimulation of lymphocytes. Cytometry, 1991. 12(1): p. 42-9.
587. Himmel, M.E., K.G. Macdonald, R.V. Garcia, et al., Helios+ and Helios- Cells Coexist within the Natural FOXP3+ T Regulatory Cell Subset in Humans. J Immunol, 2013. 190(5): p. 2001-8.
588. Huarte, M., M. Guttman, D. Feldser, et al., A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell, 2010. 142(3): p. 409-19.
589. Apparatus, B.H.; Available from: http://www.cytopulse.com/pulseagile.html.
590. Estornes, Y., F. Toscano, F. Virard, et al., dsRNA induces apoptosis through an atypical death complex associating TLR3 to caspase-8. Cell Death Differ, 2012. 19(9): p. 1482-94.
591. Walker, L.S., A. Chodos, M. Eggena, et al., Antigen-dependent proliferation of

CD4+ CD25+ regulatory T cells in vivo. J Exp Med, 2003. 198(2): p. 249-58.
592. Walker, L.S., CD4+ CD25+ Treg: divide and rule? Immunology, 2004. 111(2): p. 129-37.
593. White, J. and S. Dalton, Cell cycle control of embryonic stem cells. Stem Cell Rev, 2005. 1(2): p. 131-8.
594. Liu, G., K. Yang, S. Burns, et al., The S1P(1)-mTOR axis directs the reciprocal differentiation of T(H)1 and T(reg) cells. Nat Immunol, 2010. 11(11): p. 1047-56.
595. Canavan, J.B., B. Afzali, C. Scotta, et al., A rapid diagnostic test for human regulatory T-cell function to enable regulatory T-cell therapy. Blood, 2012. 119(8): p. e57-66.
596. Doyle, M., L. Badertscher, L. Jaskiewicz, et al., The double-stranded RNA binding domain of human Dicer functions as a nuclear localization signal. RNA, 2013. 19(9): p. 1238-52.
597. White, E., M. Schlackow, K. Kamieniarz-Gdula, et al., Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA. Nat Struct Mol Biol, 2014. 21(6): p. 552-9.
598. Mali, P., K.M. Esvelt, and G.M. Church, Cas9 as a versatile tool for engineering biology. Nat Methods, 2013. 10(10): p. 957-63.
599. Sander, J.D. and J.K. Joung, CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol, 2014. 32(4): p. 347-55.
600. Zhang, L., Q. Jiang, G. Li, et al., Efficient infection, activation, and impairment of pDCs in the BM and peripheral lymphoid organs during early HIV-1 infection in humanized rag2(-)/(-)gamma C(-)/(-) mice in vivo. Blood, 2011. 117(23): p. 6184-92.
601. Aguilera, A. and T. Garcia-Muse, R loops: from transcription byproducts to threats to genome stability. Mol Cell, 2012. 46(2): p. 115-24.
602. Ginno, P.A., P.L. Lott, H.C. Christensen, et al., R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell, 2012. 45(6): p. 814-25.
603. Skourti-Stathaki, K. and N.J. Proudfoot, A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev, 2014. 28(13): p. 1384-96.
604. Bose, P., G.L. Simmons, and S. Grant, Cyclin-dependent kinase inhibitor therapy for hematologic malignancies. Expert Opin Investig Drugs, 2013. 22(6): p. 723-38.
605. Villicana, C., G. Cruz, and M. Zurita, The basal transcription machinery as a target for cancer therapy. Cancer Cell Int, 2014. 14(1): p. 18.
606. Seumois, G., L. Chavez, A. Gerasimova, et al., Epigenomic analysis of primary human T cells reveals enhancers associated with TH2 memory cell differentiation and asthma susceptibility. Nat Immunol, 2014. 15(8): p. 777-88.